EPIGENETIC AND GENETIC PROFILES OF RARE RENAL CANCERS

By

Amy Amelia Slater

A thesis submitted to the University of Birmingham for the
degree of DOCTOR OF PHILOSOPHY

Medical and Molecular Genetics

School of Medicine & Dentistry

The University of Birmingham

September 2015

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

This study focuses upon the characterisation of the genetic and epigenetic profiles of rare forms of sporadic renal cancers (RCC), by utilising the latest technology in genome wide methylation arrays, targeted quantitative Methylight analysis and whole exome sequencing (WES). The aim was to identify differential patterns of DNA methylation or somatic mutations that may permit distinction between different subtypes of RCC which could therefore facilitate disease prognosis or help to identify molecular pathways that could be targeted therapeutically. Illumina Infinium HumanMethylation 450K BeadChip permitted the comparison of the epigenome of the malignant chromophobe RCC and the benign renal oncocytoma. This study identified twenty-eight genes to be differentially hypermethylated in chromophobe RCC, and two differentially hypermethylated genes in renal oncocytoma showing that although both visually and pathologically similar, both tumours have a distinct methylation pattern. Comparison to the methylation data available, for clear cell and papillary renal cell carcinoma from The Cancer Genome Atlas (TCGA), identified two genes *SPG20* and *NPHP4* to be significantly methylated in chromophobe RCC. Targeted Methylight analysis of the RASSF gene family in sarcomatoid RCC identified that frequent hypermethylation of *RASSF10* was significantly associated with the occurrence of metastasis and later stages of the cancer. WES of renal oncocytoma samples identified somatic mutations in eighteen genes involved in a variety of cellular functions. Sanger sequencing was then used to confirm the mutations identified, followed by further screening by Sanger in a cohort of additional renal oncocytoma samples to identify if the somatic mutations are recurrent. Modern high throughput and quantitative techniques have permitted further characterisation of these rare renal cancers and have enabled unique insights into their molecular genetic, findings that may hopefully be of clinical benefit in the future.

<u>Dedication</u>

I would like to dedicate the thesis to the memory of my uncle Dion, for without his

support and encouragement I would never have found science.

<u>Acknowledgements</u>

First and foremost, I wish to say a big thank you to my supervisor Prof. Farida Latif. Without her support, encouragement and vision none of this would have been possible, so thank you! Thank you as well to my other secondary supervisors Prof. Eamonn Maher and Dr. Mark Morris.

I would like to thank everyone in the Medical and Molecular Genetics department past and present, for all the help, support and guidance you have provide me throughout this journey. A special thanks goes to Abdullah Alholle and Dr Mark Morris, who have so kindly stepped in, to continue my project. I would also like to thank David Matallanas and his team for hosting me in Dublin and teaching me about co-IPs and proteomics.

A special thank you goes to Thoraia Shinawi, Naomi Wake and Ghazala Begum, my office mates and friends who I could always turn to and who filled my PhD with laughs and pranks, with an extra special thanks to Diana Walsh. I would also like to thank Abdullah again for the countless lifts he gave me, I was especially grateful when it was raining. A huge thank you also to Dr Hanne Gadeberg, and Pam and Colin Hickey, your support and proof reading skills have been invaluable to me. I cannot thank you enough. Neil Martin, you are a truly amazing friend and lifesaver! I cannot thank you enough for letting me stay when our house had sold. Finally, I would like to thank my family and most importantly my husband Matthew, for all your support and understanding throughout this whole process. You are my rock, and I am hugely appreciative for your excel skills. I thank you for your patience while I completed writing this thesis, although the same gratitude cannot be extended to the dog; she was not patient.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

27K array	Infinium HumanMethylation 27 BeadChip
3'UTR	3' untranslated region
450K	Infinium HumanMethylation 450 BeadChip methylation array
5-Aza	5-aza-2'deoxycitidne
5'UTR	5' untranslated region
5hmC	5- hydroxymethylcytosine
BHD	Birt–Hogg–Dubé
ccRCC	Clear cell renal cell carcinoma
cDNA	complementary DNA
CGI	CpG island
Chr	Chromosome
chRCC	Chromophobe renal cell carcinoma
CIMP	CpG island methylator phenotype
cN	Cancer Normals
co-IP	co- immunoprecipitation
CoBRA	Combined bisulfite restriction analysis
CpG	Cytosine-guanine di-nucleotides
DAVID	Data for annotation, visualisation and integrated discovery
ddPCR	droplet digital PCR
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
ECL	Enhanced chemiluminescence
EVS	Exome variant server
ExAC	The exome aggregation consortium
FD	Fold difference
FDR	False detection rate
FLCN	Folliculin
GOF	Gain of function
GOI	Gene of interest
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HGF	Hepatocyte growth factor receptor

HLRCC	hereditary leiomyoma renal cell carcinoma
HMT	Histone methyltransferase
HPRC	hereditary papillary renal cancer type 1
IGV	Integrative genomics viewer
IP	immunoprecipitation
IPA	Ingenuity pathway analysis
KD	Knockdown
LFQ	Label free quantification intensity
LOF	Loss of function
LOH	Loss of heterozygosity
MAF	Minor allele frequency
MI	Methylation index
miRNA	microRNA
mRNA	messenger RNA
MS	Mass Spectrometry
MSP	Methylation specific PCR
N	Normal
NGS	Next generation sequencing
NICE	National Institute for Health and Care Excellence
Noco	Nocodazole
Oc	Oncocytoma
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PMR	Percent of methylated reference
pRCC	Papillary renal cell carcinoma
PVDF	Polyvinylidene fluoride membrane
q-MSP	quantitative-MSP
RCC	Renal cell carcinoma
RNA	Ribonucleic acid
RT-PCR	Reverse transcription PCR
SC	Single colony
SDS	Sodium dodecyl sulphate
SIFT	Sorting tolerant from intolerant

siRNA	short interfering RNAs
SNP	single nucleotide polymorphisms
SNPs	Single nucleotide polymorphisms
SNV	Single nucleotide variant
sRCC	Sarcomatoid renal cell carcinoma
T	Tumour
T1	Thymidine 1x
TCGA	The Cancer Genome Atlas
Thy	Thymidine
TKI	Tyrosine-kinase inhibitor
TSGs	Tumour suppressor genes
TSS	Transcription start site
USP	Un-methylation specific PCR
VHL	Von Hippel–Lindau
WES	Whole exome sequencing
WGA	Whole genome amplification
WGBS	Whole genome bisulfite sequencing
WGS	Whole genome sequencing
wt	Wild type

<h1 style="text-align:center;">Chapter One: **Introduction**</h1>

1.1 <u>Cancer</u>

Cancer is the broad term given to a collection of diseases, characterised by unregulated cell growth that can spread and infiltrate surrounding tissues (King, 2006; National Cancer Institute, 2015).

Statistically, cancer is the world's leading cause of mortality, associated with over 8.2 million reported deaths in 2012 alone (Stewart and Wild, 2014) and reported as the leading cause of death in the UK, Australia, New Zealand, Canada and Denmark (Wishart, 2015).

It is predicted that there are 32 million people world-wide currently living with cancer (Ferlay et al., 2014). Within the UK, over 290,000 cancer diagnoses were registered in 2013; according to previously reported figures, this corresponds to an increase in incidence rate of 2.8% for males, and 8.3% for females over a course of ten years (Office for National Statistics, 2015a). However, it should be noted that the number of people surviving cancer is also increasing, current estimates predict that over 50% of people diagnosed with cancer survive the disease (Quaresma et al., 2015; Wishart, 2015).

It is also predicted that at least 1/3rd of cancers are preventable, with many environmental stimuli promoting cancer formation. For example 42% of cancers diagnosed in the UK are reported to be a consequence of lifestyle and environmental risk factors (Parkin et al., 2011). One of the most recognised carcinogenic risks is tobacco smoking, which has been associated with a multitude of cancers ranging from cancer of the throat to cervical and kidney cancers. In the UK, smoking is estimated to be responsible for approximately

19.4% of all newly diagnosed cancer cases (Parkin et al., 2011) and 28% of fatal cancers (in 2005) (Swerdlow et al., 2010). Other environmental factors associated with the development of cancer include; diet, obesity and high body mass index (BMI), exposure to radiation, viral infections (e.g. Hepatitis C virus and Human papilloma virus) alcohol consumption and environmental carcinogenic pollutants (Benn et al., 2016; Stewart and Wild, 2014; Swerdlow et al., 2010).

1.2 Cancer as a genetic disease

Cancer may arise from virtually any tissue or cell type, although epithelium cells are the cell type from which most common cancers arise (such as lung cancer and colorectal cancer) (Ruddon, 2007). In normal cells, cellular growth is a tightly regulated process and is a balancing act between pro-growth factors and pro-apoptotic and senescence factors. In cancer this balance is tipped in favour of growth, by the augmentation of cellular processes and the accumulation of genetic alterations (including; genetic mutations, chromosomal translocations, copy number variations (CNVs) or epigenetic augmentation) (Burrell et al., 2013).

Carcinogenesis revolves around a central principle of clonal evolution proposed in 1976 by P Nowell. He proposed that tumours originate from a single founder cell that has acquired specific genomic alterations, enabling it to escape normal cellular regulatory mechanisms (including signals for growth, respiration and apoptosis) (Nowell, 1976). Successfully avoiding these cellular regulatory signals enables uncontrolled growth and proliferation. Further genetic variations may then be acquired in the tumour that provide a cellular growth advantage; the

clones containing this advantageous variant may then undergo, subclonal selection, a process which then drives the formation and progression of tumours. Sequential growth, acquisition of genetic variations and subclonal selection drives the formation of tumours. A multi-step process resulting in normal regulated cells, becoming pre–cancerous cells then cancerous cells and finally resulting in a tumour and metastasis (Greaves and Maley, 2012; King and Robins, 2006; Nowell, 1976).

To become a cancer, a cell has to acquire a collection of ten distinct biological capacities, dubbed 'Hallmarks' (Cavallo et al., 2011; Hanahan and Weinberg, 2011). These characteristic hallmarks possessed by cancer cells are described by Hanahan and Weinberg (2011) and include: sustaining proliferative signalling, evading growth suppressors, avoiding immune destruction, resisting cell death, genome instability and mutation, inducing angiogenesis, enabling replicative immortality, activating invasion and metastasis, tumour-promoting inflammation and deregulating cellular energetics (Hanahan and Weinberg, 2011).

 The acquisition of these biological hallmarks permits the formation of the tumour microenvironment, which supports and promotes tumour growth and development. But also provides possible therapeutic targets for the treatment of cancer. For example anti VEGF molecules (such as tyrosine kinase inhibitor (TKI) sunitinib) have been approved to inhibit angiogenesis in several cancers including kidney cancer (Garajová et al., 2015; Hanahan and Weinberg, 2011; Shuch et al., 2015).

1.2.1 Intratumour heterogeneity, driver and passenger mutations

Studies that aimed to show how genetic mutations can result in cancer, have now shown that in the majority of cases, a tumour is not made up of a single clonal population of cells, as previously thought, but can actually be made up of many different clonal populations of cells that have acquired pro-tumourigenic mutations (Burrell et al., 2013; McGranahan and Swanton, 2015; Tabassum and Polyak, 2015).

Due to the genomic instability hallmark associated with cancers, as a tumour grows and proliferates, the cells acquire an increasing number of mutations (Burrell et al., 2013; Knudson, 2001). Some of these acquired mutations will have an indifferent effect within the tumour; these are known as passenger mutations (Pon and Marra, 2015; Vogelstein et al., 2013). Other mutations will provide an active growth advantage to the cell, and will promote tumour progression (these are known as driver mutations) (Pon and Marra, 2015; Tamborero et al., 2014; Vogelstein et al., 2013). Cellular clones that contain these driver mutations may undergo positive selection, thereby creating a new sub-clonal population within the tumour that drives tumour progression, microenvironment formation and is a major cause of intratumour heterogeneity and genetic diversity (Burrell et al., 2013; McGranahan and Swanton, 2015). The presence of different subclonies within one tumour type in combination with genomic instability means that, high levels of genetic diversity can be present, with each subclone presenting with a different mutation profile. For example a study by Gerlinger et al. 2014, on clear cell renal cell carcinoma, identified 49 driver mutations and 76 somatic CNVs of which 73% and 75% respectively where identified to be subclonal (Gerlinger et al., 2014). As a consequence of the mosaic nature of tumours, the original

founding mutations maybe diluted and therefore identified at low-levels, making identifying the exact causes and pathways that have been deregulated a complex and challenging task.

1.2.2 Oncogenes and tumour suppressor genes

In normally functioning cells, the genes responsible for promoting cell growth are known as proto-oncogenes, while those that prevent cell growth and proliferation and initiate apoptosis are known as tumour suppressor genes (Hanahan and Weinberg, 2011; Lee and Muller, 2010; Weinberg, 2013). When these genes become altered by intragenic mutations, chromosomal rearrangements, chromosomal duplications or deletions, for example, these proto-oncogenes and tumour suppressor genes can become oncogenic.

Proto-oncogenes were initially identified via investigations into the retrovirus transforming gene (*SCR*) and its involvement in sarcoma formation in chickens in the 1970s (Bell, 2010; Stéhelin, 1995 and references within). This discovery led to the understanding that cancer could arise through genetic mutations; a ground breaking discovery that sparked a wide interest into how genetic mutations can result in cancer formation. Since then, a large amount of information has been gained into the genetic processes behind tumourigenesis. For example, it is now widely understood that a proto-oncogene can become an oncogene when a specific genetic alteration is inherited or acquired that causes the encoded protein to become permanently active or to become over expressed in the cell. This therefore results in the constitutive and unregulated activation of signals that promote cell growth and proliferation, which can then result in tumourigenesis (Lee and Muller, 2010; Thomas et al., 2007; Weinberg, 2013).

Since the identification of proto-oncogenes, their mechanisms of activation during tumourigenesis and the significance of their role in cancer has been widely studied. One such example is the *RAS* family of GTPases; some of the most well recognised oncogenes The *RAS* genes encode small GTPases that are involved in initiating and propagating signalling cascades that result in cell growth. The RAS GTPases function by cycling between an active GTP-bound state and an inactive GDP-bound state. This cycling occurs in the absence of external stimuli as the RAS protein has the intrinsic ability to perform GTP-hydrolysis, enabling it to regulate itself. For this reason, the presence of a specific genetic mutation that affects any of the amino acid residues responsible for GTP-hydrolysis can inhibit the protein's ability to cycle to the inactive GDP-bound state. This therefore causes the RAS protein to remain in a constitutively active GTP-bound state, resulting in a permanently active signalling cascade promoting growth and proliferation, which ultimately leads to tumour formation (Pylayeva-Gupta et al., 2011). This method of tumour genesis is relatively common, for example it is predicted that >20% of cancers harbour an activating mutation for either *KRAS, HRAS* or *NRAS,* with the majority of GOF mutations located in 1 of three key amino acids *Glycine-12 (G12), Glycine-13 (G13) and Glutamine-61 (Q61)* which are essential for GTP-hydrolysis (Boichard et al., 2012; Frattini et al., 2004; Gripp and Lin, 2012; Karakas et al., 2006; Rajalingam et al., 2007)

1.2.2.1 *Tumour suppressor genes (TSGs)*

TSGs play a vital protective role in cells and are often involved in processes such as repressing cell growth, promoting cell cycle arrest, initiating DNA damage repair and inducing apoptosis (Hanahan and Weinberg, 2011; Lee and Muller,

2010; Weinberg, 2013). For this reason loss of function or reduced expression of these genes can result in unregulated and uninhibited cellular growth, proliferation and survival, thereby resulting in tumourigenesis (Hanahan and Weinberg, 2011). A well known example of this can be found in individuals who harbour mutations in either of the *BRCA1* or *BRCA2* genes. Pathogenic mutations in either of these genes can increase an individual's predisposition to breast and ovarian cancer by approximately 88% (Antoniou et al., 2004). This is because *BRCA1* and *BRCA2* are tumour suppressor genes involved in regulating DNA-damage repair; therefore loss of function of either of these genes deprives the cell of it's ability to repair damaged DNA, which results in an increase in genomic instability and the formation of a tumour as a consequence of this (Antoniou et al., 2004).

1.2.2.2 Knudson two hit hypothesis (TSGs)

Unlike oncogenes, where often only a single mutation, or 'hit', is sufficient to result in increased protein activity, TSGs often require two hits resulting in the loss of protein function, due to the fact that both alleles must be altered. This phenomena was originally described by Alfred Knudson in 1971, when he proposed the 'two-hit hypothesis' following his investigation into sporadic and inherited retinoblastoma, (a childhood cancer of the retina). Knudson observed that inheriting a single germline mutation in *RB1,* did not appear to be sufficient to cause the cancer in the offspring of parents that had previously presented with retinoblastoma. Knudson also made the observation regarding the age of the children that had inherited a germline mutation in *RB1* who then developed retinoblastoma often did so at a much younger age than those children without a

familial link to the disease. He also noticed that the affected children with a family history of the disease often developed bilateral retinoblastoma in contrast to the, often, unilateral development of retinoblastoma of the children without family histories. He therefore, deduced that the development of retinoblastoma must be a bilateral event, requiring two mutations (one in each allele). Concluding that those affected children with a familial link to the disease must have inherited one mutation (hit one), while; the other must have been acquired somatically later in life (hit two). He suggested that the affected children that had not inherited the first hit, must have acquired both mutations somatically during life, which is why they developed cancer at a later age than the inherited cases (Knudson, 1971). In 1987, *RB1* became the first TSG to be isolated, cloned and sequenced (Lee et al., 1987). These findings laid the foundations of what we now understand to be tumour suppressor genes, and initiated a new wave of investigations into how genetic mutations in either proto-oncogenes or tumour suppressor genes can result in cancer.

1.3 <u>Epigenetics</u>

Epigenetics is widely regarded as heritable alterations in DNA expression that do not alter the DNA sequence. Usually achieved by modifications to the DNA and chromatin, which alters or regulates the binding of transcription factors and other regulatory complexes and is essential form normal cell growth (Gibney and Nolan, 2010). Such epigenetic mechanisms identified, include histone modifications, DNA methylation and small non-coding RNAs such as microRNAs (miRNAs) and short interfering RNAs (siRNAs), and large non-coding RNA (Gibney and Nolan, 2010).

1.3.1 Histone modifications

Histones are highly conserved globular proteins, which form nucleosomes, vital for the packaging of DNA into chromatin. Posttranslational modifications to the N terminal tails of the histone proteins (histone tails) regulate the dynamic processes of transcription, DNA repair and replication, by altering either the chromatin density or transcription factor and other proteins accessibility and binding to the DNA (Gibney and Nolan, 2010; Kouzarides, 2007). Nucleosomes form a key component of chromatin. Each nucleosome is comprised of an octamer core of histones (histone 3,4 2A and 2B), around which 147bp of DNA is wrapped, the N terminal tail of each histone protein projects from the core and is the target for posttranslational modifications (Figure 1-1-1A) (Gibney and Nolan, 2010; Kouzarides, 2007; Venkatesh and Workman, 2015)

Organisation of DNA into chromatin and then chromosomes is achieved through several levels of organisation, starting with the DNA being wrapped around the nucleosomes, the analogy 'beads on a string' is often used to describe the appearance of chromatin at this level of organisation. These strings are then folded into a chromatin fibre of 30nm diameter by the addition of linker histone H1. Followed by further packaging and folding is undertaken to achieve the higher order of structure observed in a mitotic chromosome (Figure 1-1 -1B) (Baylin and Jones, 2011; Felsenfeld and Groudine, 2003; Venkatesh and Workman, 2015).
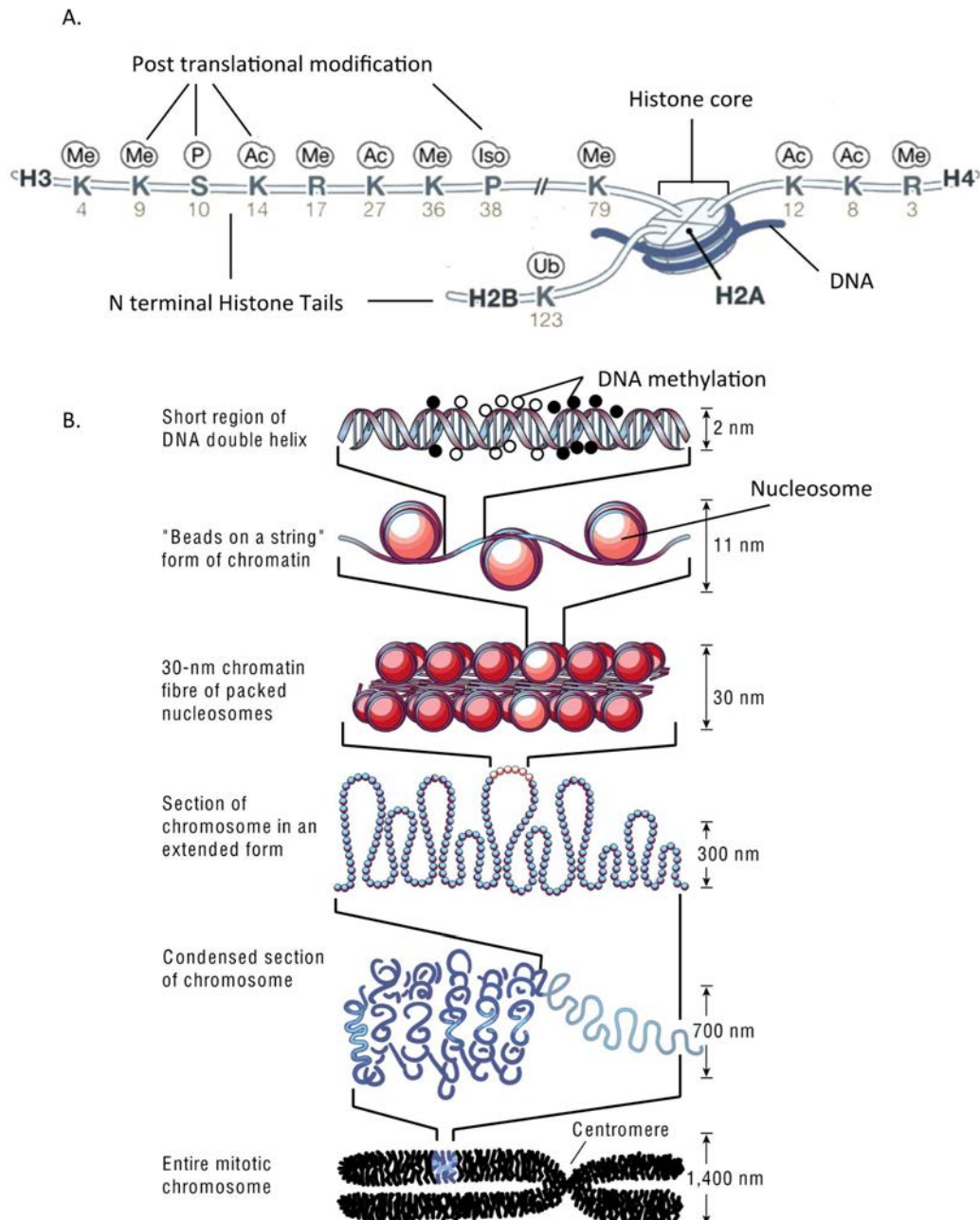
Figure 1-1: Depiction of histones structure and chromatin organisation
A) Representation of a nucleosome, depicting the histone core, histone tails and the locations of common post translational modifications (Me: Methylation Ac: Acetylation, P: Phosphorylation and Ub: Ubiquitination) (adapted from (Kouzarides, 2007). **B**) The levels of chromatin organisation within cells. Epigenetic regulation via DNA methylation occurs at the DNA level, while histone modifications affect the nucleosomes (as depicted in figure 1a) at the first level of organisation dubbed 'Beads on a string'. Adapted from (Felsenfeld and Groudine, 2003)

The post-translational modifications of the nucleosomes histone tails can include acetylation, methylation and ubiquitylation of lysine residues, methylation of arginine residues and phosphorylation of serine or threonine (Kouzarides, 2007; Sawan and Herceg, 2010). These modifications are achieved through a myriad of enzymes, including; histone acetyltransferases (HATs) (such as Hat1, Hpa2 and P300), histone deacetylases (HDACs) (e.g. SirT2 and HDAC10), histone methyltransferase (such as the arginine methyltransferase PRMT4, or lysine methyltransferases such as the MLL family or SET1A), Serine/Threonine kinases (including MSK1/2 and Mst1) and Ubiquitilases (including Ring1A/B) (Ceccacci and Minucci, 2016; Kouzarides, 2007; Sawan and Herceg, 2010).

Different histone modifications have been associated with different biological functions, for example acetylation of lysine residues within the histone tail is associated with increased gene expression via relaxation of chromatin structure. It is reported that, the addition of an acetyl group to a lysine residue, neutralises the positive charge of the histone tail, thus slightly reducing the strength of attraction to the negatively charged DNA and thereby weakening the nucleosome-DNA binding and nucleosome-nucleosome interaction. This in turn results in reduced chromatin stability, and permits transcription factors access to previously 'locked away' genes. (Baylin and Jones, 2011; Kouzarides, 2007; Rogenhofer et al., 2012b; Sawan and Herceg, 2010).

Histone methylation on the other hand has a diverse range of functions, depending upon the level of methylation and which residue is methylated, including protein recruitment and regulating chromatin density, (Sawan and Herceg, 2010). For example, histone 3 lysine 4 methylation (H3K4m) is

predominantly present at the promoter region of actively transcribed genes (Gibney and Nolan, 2010). Additionally it has been shown that during foetal development H3K4 methylation maintains gene transcription at active promoter regions by inhibiting the binding of DNA methyltransferase 3 (DNMT3) and 3L (DNMT3L), preventing methylation of the DNA and subsequent gene inactivation (Cedar and Bergman, 2009). Furthermore methylation of histone 3 lysine 36 (H3K36) and histone H3 lysine 79 (H3K79) are also associated with transcriptional activation and gene expression, with H3K36 tri-methylation enriched within the gene bodies (particularly the exons) of transcriptionally active genes (Gibney and Nolan, 2010; Sawan and Herceg, 2010; Venkatesh and Workman, 2015). Histone methylation modifications associated with transcriptional repression and heterochromatin formation include histone 3 lysine 9 (H3K9), histone 4 lysine 20 (H4K20) and histone 3 lysine 27 (H3K27) (Sawan and Herceg, 2010). With H3K27 methylation associated with the recruitment of the chromodomain containing polycomb protein PC2, a protein associated with the maintenance of X- inactivation (Kouzarides, 2007).

To add further complexity, histones tails can be methylated at three different levels, mono (me1), di (me2) and tri (me3) methylation. Conventionally the increasing number of methyl groups correlates with a more intense response, however this is not always the case, for example it is reported that mono-methylation of histone 3 lysine 27 (H3K27me1) and histone 3 lysine 9 (H3K9me1) is associated with transcriptional activation, while di- and tri - methylation (me2 and me3) of the same residues is associated with transcriptional repression (Ellinger et al., 2010; Rogenhofer et al., 2012a, 2012b). Histone modifications have a pronounced effect on chromatin stability

and gene expression and even regulate key biological features such as DNA repair and DNA replication (Baylin and Jones, 2011), it is therefore unsurprising that histone modifications are frequently associated with cancer (section 1.3.4)

1.3.2 DNA methylation

DNA methylation, is also associated with gene repression, regulation of the chromatin architecture, tissues specific gene silencing and X-inactivation (Brena and Plass, 2009; Sharma et al., 2010)and almost exclusively occurs at cytosine-guanine (CpG) di-nucleotides (Mikeska and Craig, 2014).

It is estimated that between 70-80% of all CpG loci are methylated. The majority of CpG loci are located in CpG rich regions known as CpG islands, the surrounding less CpG dense areas are referred to as shores and shelves. CpG islands are characterised as regions with >50% CpG content. The shore regions are less dense than the CpG islands and located within 2kb up or downstream of a CpG island, while CpG shelves are located >4kb up or downstream of the CpG island (Dedeurwaerder et al., 2011; Malzkorn et al., 2011; Rodríguez-Paredes and Esteller, 2011). Furthermore, a predicted 60% of gene promoter regions are believed to be regulated by CpG islands with reports also suggesting that methylation of the CpG shores too play an important role in the regulation of gene expression (Irizarry et al., 2009; Rodríguez-Paredes and Esteller, 2011). Additionally, intragenic DNA methylation is frequently observed within gene bodies, particularly within exonic regions. It is suggested that the presence of this intragenic methylation prevents random activation of gene transcription at a location that is not the proper transcription start site (TSS), promotes elongated transcription to ensure complete transcription of all exons and may even be

involved in post transcriptional splicing of the RNA (Jones, 2012). Furthermore, DNA methylation is also observed at highly repetitive intergenic elements, and is reported to promote chromosome stability by silencing transposable DNA elements (Jones, 2012; Karami et al., 2015; Sharma et al., 2010).

During methylation the cytosine residue is converted to a 5-methylcytosine (5mC) by the covalent addition of a methyl group to the carbon-5 atom. The presence of a methyl group is reported to project into the major grove of the DNA, thus repressing gene transcription through inhibiting the binding of transcription factors such as CTCF or c-myc, (Gibney and Nolan, 2010; Jones, 2012; Mikeska and Craig, 2014). Furthermore, there is a close link between DNA methylation and histone modifications, regarding regulation of gene expression and chromatin structure. Not only can DNA methylation provide binding sites for other proteins such as HDACs, resulting in histone modification and gene repressions (Sharma et al., 2010), histone modifications such as H3K9me2 have been reported to prevent DNA methylation from de-methylation ensuring prolonged gene silencing. This is achieved by the recruitment of Uhrf1, a protein which in turn recruits DNA methyltransferase 1 (DNMT1) and histone methyltransferase to maintain the methylation status of DNA and H3K9 respectively (Cedar and Bergman, 2009; Rose and Klose, 2014). DNA methylation has also been linked to the repackaging of chromatin after mitosis. During the cell cycle histones are removed to permit replication of the DNA, while DNA methylation is maintained. Once the cell has divided the histones are then replaced. The DNA methylation provides binding sites for the histone modification enzymes and acts as reference for which regions of the DNA are

repressed (high DNA methylation) and should be packaged as heterochromatin (Cedar and Bergman, 2009).

Co-ordination of DNA methylation is regulated by a family of DNA methyltransferases (DNMT) enzymes. Of note, are DNMT 3a and 3b, which are involved in *denovo* DNA methylation (Gibney and Nolan, 2010; Jones, 2012), and DNMT1, which is involved in maintaining the DNA methylation, and reproducing DNA methylation pattern during cell cycle, by replicating the methylation pattern of hemi-methylated DNA on to the newly synthesised DNA molecules during the cell cycle, thus ensuring heritability of the DNA methylation profile throughout cell passages (Avissar-Whiting et al., 2011; Heyn et al., 2012).

1.3.3 Other epigenetic regulation

Further epigenetic mechanisms involved in the regulation of gene expression include some non-coding small RNA molecules, such as microRNAs (miRNAs) or small interfering RNS (siRNA). These are short fragments of RNA (approximately 22 nucleotides in length) (Gibney and Nolan, 2010), which bind to complementary messenger RNA (mRNA) and either inhibit translation or prompt cleavage or degradation of the mRNA, thus inhibiting translation of the mRNA into protein. The mechanism by which miRNAs enforce gene repression is dependent on how complementary the miRNA is to the mRNA, perfect complementary binding, targets the mRNA for degradation, while less than perfect complementary binding simply inhibits ribosomes binding (Gibney and Nolan, 2010; Hoffman and Cairns, 2011; Sato et al., 2011).

Most studies into the epigenetics of small noncoding RNAs, focus on miRNA. miRNA is transcribed as a pre-miRNA hairpin loop, which is then processed via

the proteins Drosha and Dicer into mature miRNA. Where it then incorporates with the RNA-induced silencer complex (RISC) to bind and regulate mRNA (Sato et al., 2011). Over 1000 miRNAs have so far been identified, many of which reside within intergenic regions of the genome with expression under the regulation of DNA methylation (Gibney and Nolan, 2010; Lopez-Serra and Esteller, 2012; Sato et al., 2011). The deregulation of miRNAs has been associated with several diseases including cardiovascular disease and cancer (Chuang and Jones, 2007; Hoffman and Cairns, 2011; Sato et al., 2011) while miRNA expression profile have also been investigated as a method to differentiate between different cancers types (Chuang and Jones, 2007; Youssef et al., 2011).

Another epigenetic marker recently identified is 5-hydroxymethylcytosine (5hmc), generated by the removal of DNA methylation via TET enzyme (Roy et al., 2014; Sajadian et al., 2015). However, to date little is known about the 5hmC biological function, but it is predicted to play a role in DNA methylation, with loss of 5hmC having been associated with melanoma and the progression of hepatocellular cancer (Lian et al., 2012; Sajadian et al., 2015). However, analysis of 5hmC presents a challenge as conventional bisulfite modification techniques cannot distinguish between 5mC and 5hmC. Currently analysis therefore requires immunoprecipitation of DNA utilising 5hmC specific antibodies (Jin et al., 2010; Sajadian et al., 2015).

1.3.4 Epigenetics in cancer

Epigenetic deregulation is widely associated with and studied in many different cancers. With aberrant epigenetic profiles identified as one of the mechanisms

contributing to the Knudson's two-hit hypothesis for the loss of heterozygosity (Waldmann and Schneider, 2013).

As introduced above, the epigenome is a complex system regulated by many factors and enzymes, deregulation at any stage may contribute to the formation of cancer hallmarks.

Studies into the histone modification patterns of different cancers have identified frequent deregulation of the enzymes and co-factors that regulate the post-translation histone modifications (Avvakumov and Côté, 2007; Cao and Zhang, 2004; Dalgliesh et al., 2010; Feinberg, 2007), leading to abnormal chromatin packaging, gene expression and DNA methylation profiles (TCGA, 2013). For example, overexpression of histone methyltransferase EZH2 has been associated with poor prognosis in breast and prostate cancer, and is predicted to result in widespread transcriptional silencing by widespread methylation of H3 lysine-27 (H3K27) (Cao and Zhang, 2004; Feinberg, 2007). Somatic mutations of the *UTX* gene, in several cancer types, including; colorectal adenocarcinoma, multiple myeloma, and oesophageal squamous cell carcinoma, results in decreased expression of the H3K27 de-methylase enzyme UTX, leading to increased H3K27 methylation and wide spread transcriptional repression (van Haaften et al., 2009).

Furthermore, deregulation and mutations of components of chromatin remodelling complex's such as the ATP-dependent SWI/SNF complex, (a complex associated with opening condensed chromatin to permit transcriptional activation) have been associated with abnormal nucleosome positioning, chromosomal rearrangements and alter gene expression in several cancers (Brock et al., 2007; Feinberg, 2007; Ferraro, 2016). For example *INI1*, *SMARCA4*,

SMARCAD1, ARID1A and *PBRM1* have previously been reported to be frequently mutated or down regulated in various cancers, such as colorectal, kidney, ovarian, rhabdoid medulloblastoma and breast cancer, resulting in chromosomal instability, abnormal gene expression and subsequent tumourigenesis (Ferraro, 2016; Pancione et al., 2013; Roy et al., 2014; Varela et al., 2011; Versteege et al., 1998; Wilson and Roberts, 2011).

INI1 (also known as *SMARCB1)* is a core component of the SWI/SNF complex, involved in DNA binding and chromatin stabilisation; loss of function mutations of *INI1* have been associated with aggressive paediatric rhabdoid tumour and poor prognosis in colorectal cancer (Pancione et al., 2013; Sévenet et al., 1999; Versteege et al., 1998; Wang et al., 2016). *SMARCAD1* has been reported to be lost in colorectal and head and neck squamous cell carcinomas, loss of this gene has been reported to lead to chromosomal instability, a characteristic hallmark of cancer (Berg et al., 2010; Cetin et al., 2008; Hanahan and Weinberg, 2011).

Similarly mutations or alteration in expression of DNA methylation regulators such as the DNMT family or the demethylation associated DNA hydroxylase enzyme *TET* are relatively frequent in cancers resulting in an abnormal global DNA methylation pattern, and have been associated with genomic instability and augmented gene expression (Heyn et al., 2012; Roy et al., 2014). For example; *DNMT1, 3A* and *3B* are frequently overexpressed in colon, prostate, leukaemia and breast cancer (Subramaniam et al., 2014) while *DNMT3A* is frequently mutated in acute myeloid leukaemia (Roy et al., 2014) and TET is frequently mutated in multiple myeloid neoplasms (Roy et al., 2014).

An initial study on DNA methylation in cancers by Gama-Sosa et al. in 1983, reported hypomethylation at many promoter regions associated with the

activation of gene transcription (Gama-Sosa et al., 1983). Subsequent investigations into promoter hypomethylation focused on proto-oncogenes, such as *HOX11*. A proto-oncogene identified to be hypomethylated in T-cell acute lymphoblastic leukaemia, resulting in increased *HOX11* expression and cell growth (Ehrlich, 2002; Watt et al., 2000). Furthermore global hypomethylation has been observed in several cancer types, within intergenic regions of DNA and at repetitive elements(Ehrlich, 2009) and is believed to be involved in the relaxation of heterochromatin, genomic instability and activation of repetitive transposable elements such as LINE-1, resulting in DNA arrangements and gene disruption (Ehrlich, 2009; Lorincz et al., 2004). Recent investigations, suggest that the presence of global hypomethylation may be a biomarker for early stage cancer, having been identified in pre-cancerous breast tissue and early stage urothelial cancer (van Veldhoven et al., 2015; Wolff et al., 2010), allowing rapid intervention.

However, the majority of studies on DNA methylation in cancers have primarily focused on hypermethylation and its role in silencing proteins encoded by TSG. Cancer-specific DNA hypermethylation has been reported to predominantly occur within CpG islands (Costello et al., 2000). It is currently predicted that between 5 – 10% of gene promoters are hypermethylated in cancer. This suggests that, in these cases, the mechanism of tumourigenesis likely involves the inhibited transcription of the associated genes, resulting in the repression of an array of downstream signalling pathways, ultimately resulting in unregulated and excessive cell growth (Figure 1-2) (Baylin and Jones, 2011; Esteller et al., 2000; Heyn and Esteller, 2012; Lopez-Serra and Esteller, 2012; Mizuno et al., 2002).
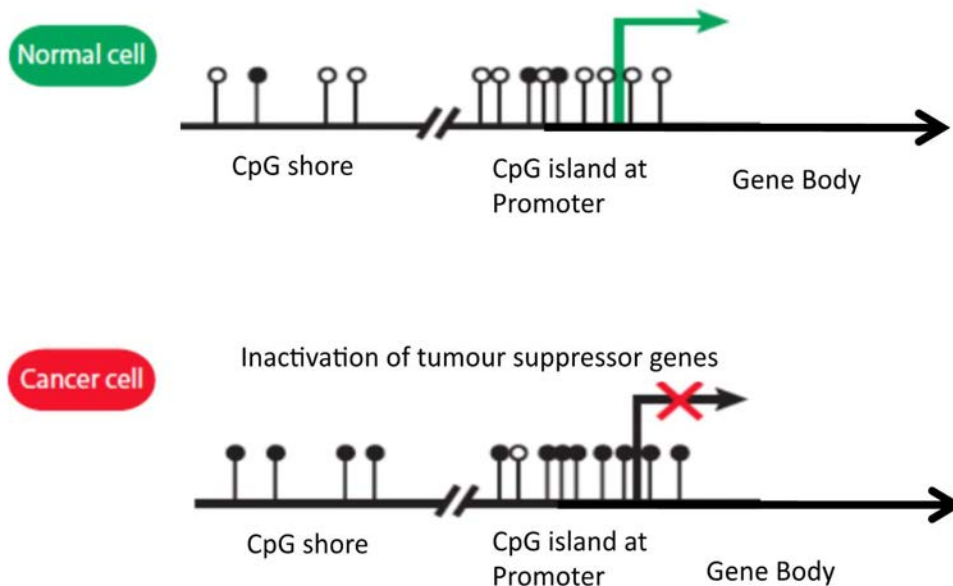
Figure 1-2: Schematic of DNA methylation at CpG island of a TSG
Visualisation of the effect of DNA methylation on the transcription of TSG in cancer cells compared to normal cells. Each circle represents a CpG locus. In Normal cells, CpG islands are unmethylated (white circles), permitting the binding of transcription factors and the expression of the gene (green arrow). In cancer, CpG loci of TSG are methylated (black), inhibiting transcription of the gene. Image from Malzkorn et al., 2011.

The study of DNA methylation in cancers not only provides invaluable insights into the mechanisms of tumour development, but can also be used as a biomarker, which can be used to distinguish between tumour types, which therefore facilitate accurate diagnoses. For example, the frequent levels of methylation of *MGMT, RASSF2* and *WIF1* have been proposed as biomarkers for diagnosis with high sensitivity and specificity for colorectal cancer (Heyn and Esteller, 2012).

Furthermore, DNA methylation status within a tumour has also been reported to be useful as a biomarker to provide an indication of prognosis, and can also be used to predict how these tumours might respond to various treatments (Heyn et al., 2013; Strimbu and Tavel, 2010). For example Exner et al 2015, reported that the presence of hypermethylation of *CDKN2A* shows a strong correlation with poorer prognoses and lower survival rates in colorectal cancer (Exner et al., 2015). Other studies have found that *CDKN2A* methylation can act as a marker to indicate early stage adenocarcinoma, while other genes such as SFRP1 (a WNT pathway inhibitor) usually acquire methylation in the more advanced stages of cancer (Selamat et al., 2011).

Another example of the diverse amount of information that can be gained by observing the DNA methylation status of a tumour is the associations between DNA methylation and the cancer's response to a particular therapeutic treatment. For example, it was recently found that hypermethylation of the *MGMT* gene in glioblastoma patients was associated with a better clinical response for patients receiving the chemotherapy treatment, temozolomide, in comparison to those patients who did not exhibit MGMT hypermethylation

(Heyn and Esteller, 2012; Trabelsi et al., 2015). Suggesting that DNA methylation may also have an effect on the efficiency of drugs.

As previously mentioned miRNAs have also been implicated in cancer development. Many studies have been undertaken to characterise the miRNA expression profile in different cancers and ascertain if there is an association between miRNA profile and cancer type (Chuang and Jones, 2007; Youssef et al., 2011). Studies have identified that depending on the target of the miRNA, they may posses an oncogenic or tumour suppressive role (Chuang and Jones, 2007). The disruption of the expression of miRNAs, by, mutations, abnormal DNA methylation of the miRNA resulting in over or under expression or genomic instability resulting in copy number alterations (Lopez-Serra and Esteller, 2012) is often attributed to cancer development. For example miR-155 has been reported to be overexpressed in breast, colon and B-Cell lymphomas and associated with oncogenic development (Chuang and Jones, 2007), while miRNA *let7* which targets oncogene RAS has been reported to be down regulated in lung cancer leading to tumour genesis (Chuang and Jones, 2007; Sato et al., 2011)

The prevalence of epigenetic deregulation in cancer has led to investigations targeting epigenetic modifications therapeutically, with several drugs obtaining approval for clinical use (Azad et al., 2013). For example suberoylanilide hydroxamic acid was the first histone deacetylase inhibitor approved for the treatment of T cell cutaneous lymphoma (Gal-Yam et al., 2008). While DNMT inhibitors Decitabine and 5-azacytidine, have been approved for treating bone marrow disorders such as myelodysplastic syndrome and acute myeloid leukaemia (Subramaniam et al., 2014). With further clinical trials on going to assess other DNMT inhibitors in cancers including ovarian, lymphomas and non-

small cell lung cancer (Subramaniam et al., 2014). However due to the ubiquitous expression of these epigenetic modifications across cell types, epigenetic targeted therapies tend to be present with high toxicity and adverse reactions (Azad et al., 2013; Gal-Yam et al., 2008; Heyn and Esteller, 2012; Peedicayil, 2006; Subramaniam et al., 2014).

1.3.4.1 *CpG island methylator phenotype (CIMP)*

The term CIMP was first coined in 1999, in colorectal tumours, and refers to a subpopulation of samples with an abnormally high frequency of global hypermethylation in cancers (Toyota et al., 1999). Identified by clustering analysis of genome wide DNA methylation profiles, the CIMP pattern has been identified and is often associated with poor clinical outcomes including increased microsatellite instability (Issa, 2004; Toyota et al., 1999), poor overall prognosis (Issa, 2004), and more aggressive forms of cancer (Arai et al., 2012). However glioblastoma has been reported to be an exception to this observation, as when coupled with *IDH1* mutation CIMP positive samples were associated with good prognosis and longer term survival (Noushmehr et al., 2010; Shinawi et al., 2013).

1.4 Kidney Cancer

Recent studies have shown that kidney cancer is the ninth most common cancer globally, and is estimated to account for 2-3% of all adult cancers with over 337,000 new cases reported in 2012 alone (Escudier et al., 2012; Henrique and Jer, 2015; Jonasch et al., 2014; Maher, 2013; Stec et al., 2009). Of these cases, studies have also shown that the highest incidence of kidney cancer occurs in developed countries (Jonasch et al., 2014). The exact reason as to why kidney cancers seem to be more prevalent in developed countries remains largely

unknown, however it is possible that differences in environmental exposures, diet or lifestyle may be responsible. Several risk factors have been associated with an increased risk of developing kidney cancer. Similar to many other cancers, smoking poses one of the highest kidney cancer risks; studies have shown that smokers may have a 52-54% increased risk of developing kidney cancer risk compared to non-smokers (Hunt et al., 2005). In addition, obesity is another factor, which has been associated with the development of kidney cancer. Some studies estimated obesity to play a role in up to 24% of kidney cancers, and suggested that there may be a direct correlation between increasing Body Mass Index (BMI) and increased kidney cancer risk (Kendall et al., 2015; Parkin and Boyd, 2011). These environmental and lifestyle associations with kidney cancer susceptibility could be the reason why this cancer is said to have such a higher prevalence in developed countries. Consistently, it has also been shown that the incidence of kidney cancer is rising, in the United States (as has obesity); in fact it has been estimated that the number of kidney cancer cases diagnosed in 2014 had increased by approximately 20% in comparison to the number of cases reported in 2007 (Henrique and Jer, 2015), a staggering increase from the cases reported in 2007 (Ng et al., 2008; Vera-Badillo et al., 2012), with kidney cancer attributing to over 2% of cancer deaths in the US (Henrique and Jer, 2015; Ng et al., 2008).

Interestingly kidney cancer has been shown to display a gender bias and is more frequently diagnosed in males compared to females, and is the seventh most common cancer in males and the ninth most common cancer in females worldwide (Escudier et al., 2012; Stec et al., 2009).

Data from the National office of Statistics has identified that the incidence of kidney cancer in the UK has increased by almost a third within the last decade (Figure 1-3a). Kidney cancer was recorded to account for 3% of all new cancer diagnosis in the UK in 2014, 62.6% of these new kidney cancer diagnosis were male (37.4% were female), with the majority of kidney cancers diagnosed in the over 60s (Figure 1-3b) (Office for National Statistics, 2015b). If diagnosed at an early stage (see section 1.4.1) survival rates for kidney cancer are very promising, with a stage I RCC predicted to have a 5 year survival rate of 70- 90% (Stec et al., 2009), in the UK five year survival rate of all kidney cancers in 2014 was 57.6% for males and 60.8% for females (Office for National Statistics et al., 2015). This could be due to a number of factors, including increasing levels of obesity in the UK for example, or could be due to the presence of other environmental stimuli. Although there are many environmental carcinogens and lifestyle habits which have been linked to increased susceptibility to kidney cancer, there are also a number of genes which, when altered, have been associated with increased kidney cancer susceptibility some examples of which are described in section 1.4.2.2 Familial RCCs.

A. **Kidney Cancer (C64-C66,C68): 1993-2011**
European Age-Standardised Incidence Rates per 100,000 Population, by Sex, UK
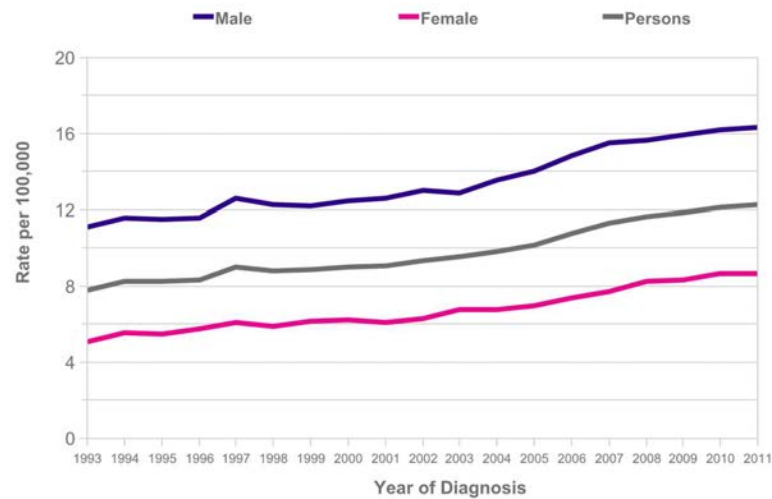
B. **Kidney Cancer (C64-C66,C68): 2009-2011**
Average Number of New Cases Per Year and Age-Specific Incidence Rates per 100,000 Population, UK
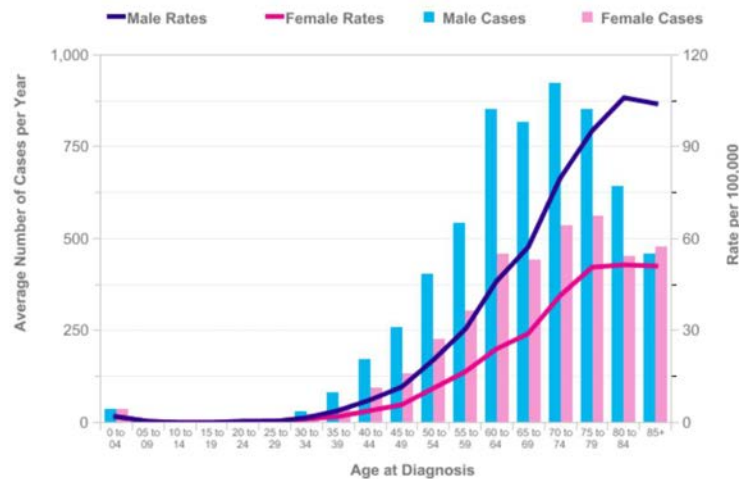
Figure 1-3: Statistic for kidney cancer in the UK
The incidence of kidney cancer is increasing for the whole population with the risk of disease greater in the older population and in males.
A) Incidence rates of all RCC diagnosis in the UK between 1993- 2001 separated by gender. **B)** Average number of new cases a year in relation to age and gender. Males over 60 years of age are those most likely to be diagnosed with a form of kidney cancer. Graphs from CancerResearchUK.org (2015).

1.4.1 Staging

Staging a cancer is a useful diagnostic tool that helps provide a more accurate disease prognosis , which may then also help to determine the most appropriate treatment options. Grading is determined through combined clinical observations and conforms to a system known as the TNM staging system (Figure 1-4) (Egner, 2010).

A summary of the TNM staging system for kidney cancer is as follows (Figure 1-4)

T refers to the tumour and is graded 0 (no evidence of primary tumour), to 4(cancer has spread past the kidney and surrounding tissues). Sub-classifications (a-c) are also documented to denote tumour size and location within the kidney, for example any tumour classified as T1 is located solely within the kidney, T1a is solely within the kidney and less that 4cm, T1b is still limited to the kidney but is between 4 and 7 cm, and T2 is greater than 7 cm but within the kidney. T3a is a tumour which extends into the adrenal gland or perinephric tissues, while a T3c tumour is classified as a tumour that extends into the vena cava, a T4 is a renal tumour which has extended beyond the connective tissue encapsulating the kidney (Egner, 2010; Ng et al., 2008).

N refers to the regional lymph nodes around the kidney, and is graded 0-2, where 0 is no cancer in any lymph nodes and 2, cancer is in more than 1 lymph node.

M is metastasis, a binary score given to whether the cancer has metastasised to another region of the body (1) or not (0).

As with most cancers, the lower the grade the better the prognosis. For RCC a tumour with a TNM score below T2 N0 M0 predicts a 5 year survival rate between 70 – 90%, while a T3a N0 M0 tumour that has invaded the perinephric fat has a survival rate of 60-80%, a T3c N0 M0 has a predicted survival rate of 0 – 30% and any RCC tumour that has metastasised (M1) has a predicted survival rate of 0 – 10% (Ng et al., 2008). Kidney cancer tumours may also be referred to as Stage I-IV (Figure 1-4), this score is calculated based upon the TNM score where:

Stage I: TNM score of T1 N0 M0

Stage II: TNM score of T2 N0 M0

Stage III: TNM scores of T1 N1 M0 to T3 N1 M0

Stage IV: is any tumour categorised as T4, or any tumour with an N score of 2 or any metastasis (Egner, 2010; Ng et al., 2008)

Currently, however, staging does not take into account other features that may impact on prognosis such as the presence of sarcomatoid cells (Humphrey, 2012) or necrosis (Brinker et al., 2000).
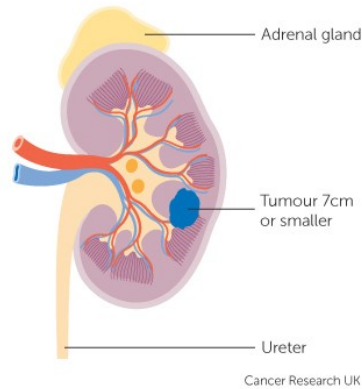
Stage 1:Tumour <7 cm diameter and is completely inside the kidney
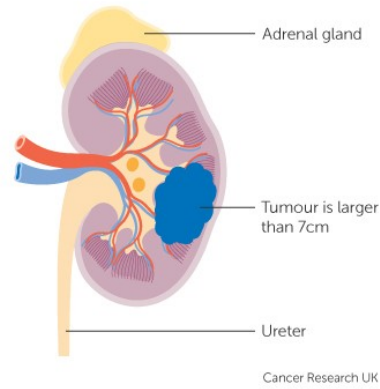
Stage 2: Tumour >7 cm diameter and is completely inside the kidney
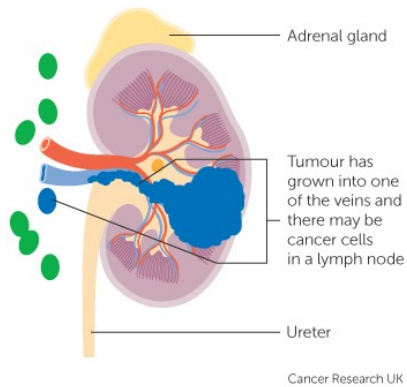
Stage 3 : Tumour has spread to adrenal gland or one of that major blood vessels. Cancer may be present in neighboring lymph nodes
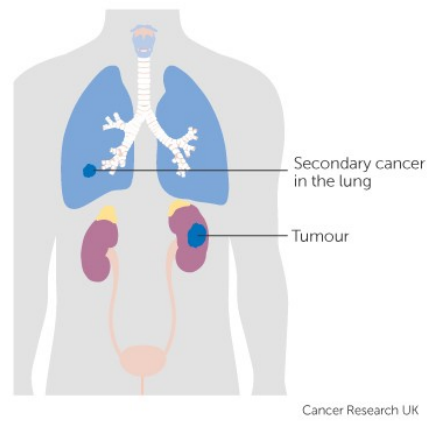
Stage 4: The cancer has metastasised to other organs ie lungs.

Figure 1-4: Visual representation of the 4 stages of kidney cancer classification.

Diagram of the 4 stages of kidney cancer classification. Stages are calculated using the TNM score.

Images adapted from CancerResearchUK.org 2015

1.4.2 Renal cell carcinoma (RCC)

Approximately 85% of all kidney cancers are classified as renal cell carcinomas (RCC) (Bodmer et al., 2002; Ng et al., 2008). RCCs are a heterogeneous group of tumours, which can be further categorised into several subtypes based upon their pathology histology, location, appearance, genetic and cytogenetic differences (Faisal et al., 1997; Kovacs et al., 1997; Linehan and Ricketts, 2014; Lopez-Beltran et al., 2009, 2006), some of these subtypes will be described in the following sections.

Generally, symptoms tend to become apparent during the later stages of RCC when the prognosis is poor. In some cases, even at very late stages, many patients are asymptomatic, and for this reason, early diagnosis is vital (Morris and Maher, 2010). Patients may present with a variety of symptoms including pain, haematuria, an abdominal mass or paraneoplastic syndromes, which include hypercalcemia, fever or wasting syndromes (Escudier et al., 2012). However, for more than 50% of cases, diagnoses are often incidental, and detected through abdominal imaging for other reasons, especially for small asymptomatic masses (Escudier et al., 2012; Jonasch et al., 2014). Diagnosis is often achieved through a combination of techniques, including medical imaging (computed tomography (CT) the most frequently used imaging technique, although ultrasound and Magnetic Resonance Imaging (MRI) is also used), in addition to blood tests and biopsies to perform pathological analysis (Escudier et al., 2012; Vera-Badillo et al., 2012). Medical imaging is currently the most favoured tool in diagnosis, as it can provide clarification on location, grade, and can even help to determine the subtype. As the different subtypes often present with distinct imaging characteristics, such as an enhanced peripheral edge of the

tumour (associated with clear cell, papillary and collecting duct RCC) or homogenous CT enhancement (associated with chromophobe and oncocytomas). However, it remains radiographically challenging to differentiate chromophobe from oncocytomas (Baert and Sartor, 2006; Millet et al., 2011; Stec et al., 2009; Vera-Badillo et al., 2012).

1.4.2.1 *Sporadic RCCs*

The majority of RCC cases are sporadic (over 96% of cases), with hereditary (familial) RCC accounting for less than 4% of all cases (Jafri et al., 2015; Lopez-Beltran et al., 2009; Maher, 2013, 2011). A summary of current genetic and histological understanding of the key sporadic subtypes is detailed below and in Table 1-1.

RCC subtype	Incidence	Development and Growth pattern	Cell/tissue characteristics	Prognosis	Genetics	
					Cyto	Molecular
Clear cell	75%	Solitary, solid tumour, cystic	clear cytoplasm; cells with eosinophilic cytoplasm seen occassionally	Aggressiveness according to grade,stage and sarcomatoid change	-3p, +5q22, +7q -6q, -8p, -9p, -14q	*VHL*, SETD2, PBRM1, JARID1C, mTOR*
Papillary	10%	Multicentric, bilateral or solitary, solid tumour, tubulo-papillary	Type 1 (basophilic) or Type 2 (eosinophilic)	Aggressiveness according to grade, stage and sarcomatoid change	+3q, +7, +8, +12, +16, +17, +20, -Y	*MET*, FH**
Chromophobe	5%	Solid Solitary tumour	Pale or eosinophilic with granular cytoplasm	10% mortality	-1, -2, -6, -10, -17, -21, hypodiploidy	*BHD*, TP53, PTEN*
Oncocytoma	5%	Solid Solitary tumour	large with granular eosinophilic cytoplasm densely populated with mitochondria	Benign	-1, -Y, 11q13 translocation	*ERCC2, C2CD4C,* mitochondrial DNA variations
Other renal carcinomas not discussed in this thesis						
Collecting ducts of Bellini	1%	Solitary tumour with irregular channels	Eosinophilic cytoplasm	Aggressive, 2/3 of patients die within two years	'-1q, -6p, -8p, -13q, -21q, -3p (rare)	-
Xp11 translocation	Rare	Solitary tumour, tubulo-papillary	Clear and eosinophilic cells	Indolent	t(X; 1) (p11.2; q21), t(X; 17) (p11.2; q25)	-
Unclassified	4%-to-6%	Solid Solitary tumour	Variable, sarcomatoid	High mortality	-	-

Table 1-1: Summery of the histological and genetic features associated with different RCC subtypes. (Adapted from Lopez-Beltran et al., 2009)

1.4.2.1.1 *Clear cell renal cell carcinoma (ccRCC)*

Amongst all classifications of RCC, clear cell renal cell carcinoma (ccRCC) is the most common, accounting for approximately 75-80% of all RCC diagnosed, of which, approximately 95% are sporadic. Although ccRCC is considered to be one of the more aggressive forms of RCC, ccRCC has a relatively high 5-year survival rate overall; up to 60%, provided that no metastasis has occurred (10% survival rate after the occurrence of metastasis) (Avissar-Whiting et al., 2011; Stec et al., 2009).

ccRCC originates from the proximal renal tubule cells and usually presents as a golden yellow solitary cortical neoplasm, which is histologically comprised of cells with clear cytoplasm with clustered eosinophilic cells in the core of the lesion (Faisal et al., 1997; Lopez-Beltran et al., 2006). The size of the tumour is variable, and other features which can be present include necrosis, sarcomatoid features, calcification and microvascular invasion; features which can all have a negative impact on the prognosis (Lopez-Beltran et al., 2009).

ccRCC has been associated with several chromosomal alterations including gain of chromosome 5q22, and loss of chromosome 6q,8p, 9p and 14q (Lopez-Beltran et al., 2009). However the most commonly observed chromosomal alteration in sporadic ccRCCs is the frequent loss of the chromosome 3p arm (over 90% ccRCC cases) (Gnarra et al., 1994; Lopez-Beltran et al., 2010), in particular, the 3p12-14, 3p21 and 3p25, often resulting in disruption of the *VHL* gene (Lopez-Beltran et al., 2010, 2009; Maher, 2013).

VHL is a well studied TSG located at 3p25-26 (Latif et al., 1993), that was originally identified as the causative gene for autosomal dominant disease von

Hippel–Lindau (VHL) (Section 1.4.2.2.1)(Gossage et al., 2015; Latif et al., 1993). The *VHL* gene encodes an E3 ubiquitin ligase tumour suppressor protein with a vital role in ubiquitinating and degrading hypoxia inducible factor-α (HIF-α) under normal conditions. Loss of *VHL* results in activation of HIF1α and HIF2. Subsequent binding to hypoxia response elements, promotes the transcription of target genes which regulate processes such as angiogenesis (Clifford and Maher, 2001). HIF signalling is further associated with the activation and signalling of the *PI3K/AKt/mTOR* pathway and the *AMPK* pathway, further highlighting the wide impact the loss of VHL can have (Gossage et al., 2015; Linehan et al., 2010). For these reasons, loss of *VHL* can therefore result in constitutively active and unregulated signals for angiogenesis and growth, among others; factors which promote tumour development and progression.

Within sporadic ccRCC *VHL* is also observed to be the most frequently mutated gene (57% of samples tested). Furthermore, in agreement with Knudson's two-hit model (section 1.2.2.2), bi-allelic loss of VHL was often reported, a result of frequent loss of heterozygosity of chromosome 3p (Gnarra et al., 1994; Lopez-Beltran et al., 2010), gene mutation or frequent DNA methylation (Section 1.5) (Herman et al., 1994).

Other genes frequently mutated, resulting in loss of function have been recently identified in sporadic ccRCC. These include; *PBRM1 (41%)* (Varela et al., 2011) *BAP1* (14%) (Peña-Llopis et al., 2012), *SETD2* (16%), *JARID1C* (4%) and *KDM6A* (3%) (Dalgliesh et al., 2010; Freier et al., 2010). Correlating with immunohistochemical studies which observed reduce protein expression ofPBRM1 and BAP1 in ccRCC (decrease of PBRM1 expression in 43% of samples

and BAP1 expression in 10% samples (Ho et al., 2015). Many of these genes are associated with chromatin regulation and remodelling; for example *PBRM1* (Polybromo-1) is a component of the chromatin remodelling complex SWI/SNF, while *SETD2* is a histone 3 lysine 36 methyltransferase, *JARID1C* is a histone 3 lysine 4 demethylase and *KDM6A* is a Histone 3 Lysine 27 demethylase.

Additionally it should be noted that *BAP1* and *PBRM1* also reside on the frequently lost chromosome 3p arm and are therefore frequently lost in ccRCC (A. Hakimi et al., 2013). However no DNA methylation has been observed for *PBRM1, BAP1, SETD2, JARID1C* or *KDMA* in ccRCC (Ibragimova et al., 2013a). Clinically, *BAP1* mutations have been associated with higher grades of ccRCC, metastatic disease and a poor prognosis and it is predicted to be a two-hit TSG (Kapur et al., 2013; S. Wang et al., 2014), while *PBRM1* and *BAP1* mutations are virtually mutually exclusive, with patients harbouring mutations in both genes rarely observed (Brugarolas, 2014; Kapur et al., 2013; Roy et al., 2014).

Recent multi-platform studies comparing the global; mutation, DNA methylation and gene expression profiles have also confirmed the observations previously mentioned; in larger cohorts (A. Hakimi et al., 2013; Sato et al., 2013; TCGA, 2013). Interestingly these larger studies have also identified significant mutations in *mTOR* (5.7-6% of samples tested), *PI3KCA* (3-4.7% of samples tested), *TP53* (~2% of samples tested) and *PTEN* (1.8% of samples tested) in addition to focal copy number deletions in *CDKN2A* 9p21 and *PTEN* 10q23. Furthermore, the studies confirmed the presence of other chromosomal abnormalities associated with ccRCC, including chromosomal gains at 5q (65%-67%) and 7q (41%), and LOH at 8p (20%), 9p (15%), and 14q (27%) (Faisal et

al., 1997; A. Hakimi et al., 2013; Sato et al., 2013; TCGA, 2013). Loss of *PTEN* and *CDKN2A* although rare, have been associated with worse overall survival in ccRCC (Feng et al., 2015). These studies highlight the wide variety of genetic alterations that can play a role in the development and progression of ccRCCs. Having a greater understanding of the roles and pathways involved in the progression of these cancers will enable the future development of more effective diagnostics, therapies and potential prophylactic treatments for these patients. For example, current *invitro* and *invivo* investigations have identified a potential PI3Kβ inhibitor which significantly reduces tumour formation when *VHL* and *SETD2* mutations were present (Feng et al., 2015).

1.4.2.1.2 *Papillary renal cell carcinoma (pRCC)*

From all types of RCC the papillary form is the second most common RCC, accounting for 10–20% all cases These forms are less aggressive compared to ccRCC, and only approximately 4% are thought to be associated with a germline predisposing mutation (Courthod et al., 2015). Derived from cells of the proximal renal tubule, pRCC is categorised into two types Type I and Type II. Although the exact proportion of each type is largely unknown, a study by Klatte at al suggested that 32% of pRCC samples are type I and 68% of pRCC samples are type II (Klatte et al., 2009), additionally a mixed histology may also be frequently observed (Yang et al., 2005).

Type I is characterised by the presence of small cuboidal cells with insufficient basophilic cytoplasm. This type is associated with a better prognosis and longer survival. *Type II* (eosinophilic) on the other hand is formed of large cells with higher nuclear grade, eosinophilic cytoplasm and pseudostratification. Type II is

associated with a poor prognosis (Faisal et al., 1997; Lopez-Beltran et al., 2006) (Faisal et al., 1997; Lopez-Beltran et al., 2006); this is likely due to the fact that type II tumours have been found to express higher numbers of receptors for vascular endothelial growth factor (VEGF) receptors, which consequently results in pro-tumourigenic effects including increased angiogenesis and micro vascularisation (Klatte et al., 2009). There are several well known alterations that are associated with pRCC; particularly chromosomal gains at 3q, 7, 8, 12, 16, 17, and 20, in addition to loss of chromosome Y with full or partial gains in chromosome 7 and 17 (which are frequent in both type 1 and type 2)(Albiges et al., 2014; Durinck et al., 2015; Faisal et al., 1997; Maher, 2013; Marsaud et al., 2015).

Several genes have been found to be associated with the hereditary forms of pRCC. For example, gain of function (GOF) mutations in *MET* are known to be frequently associated with the hereditary forms of pRCC type 1, while mutations in fumarate hydratase *FH* are associated with the inherited forms of type 2 pRCC (section 1.4.2.2.2) (Yang et al., 2005). In sporadic pRCC, GOF *MET* mutations have been reported in 15-20% of type 1 pRCC, (Albiges et al., 2014; Durinck et al., 2015), and increased copy number of *MET* has been observed in 81% of type 1 pRCC samples, showing significantly increased *MET* expression compared to ccRCC (Albiges et al., 2014; Cairns, 2010; Marsaud et al., 2015). However methylation or mutations of *FH* are seldom observed in sporadic type 2 pRCC (Cairns, 2010).

The *MET* gene encodes a hepatocyte growth factor receptor (HGF). The receptor plays a role in the activation of the PI3K pathway. When this pathway is

activated, it promotes the transcription of various nutrient transporters, which in turn may then activate the mTOR pathway, leading to cell growth and proliferation (Linehan et al., 2010). Due to its activating role in the PI3K pathway, *MET* has been found to play an oncogenic role in pRCC, with the majority of pathogenic mutations playing a gain-of-function role, resulting in constitutive protein activation and a subsequent downstream signalling cascade that results in an array of pro-tumourigenic effects, including cell growth and proliferation (Albiges et al., 2014; Durinck et al., 2015)

In addition to the genes mentioned so far, studies into sporadic pRCC have also identified significant mutations in*, NF2, (*Neurofibromatosis 2, a well known tumour-associated gene involved in the HIPPO pathway) (Evans, 2009; Harvey et al., 2013)) and *SETD2* (Durinck et al., 2015). As previously mentioned *SETD2* is known to be frequently mutated in ccRCC, (Dalgliesh et al., 2010) with mutations in *SETD2* being linked to abnormal DNA methylation profiles (TCGA, 2013)and thus may provide a pivotal role in RCC development. Interestingly, despite the known associations between *PBRM1* and ccRCC, no mutations in *PBRM1* have been identified in pRCC (Varela et al., 2011), this observation is largely collaborated by immunohistochemical staining identifying infrequent loss of *PBRM1* in 3% and *BAP1* in 0% of pRCC samples (Ho et al., 2015).

1.4.2.1.3 *Chromophobe renal cell carcinoma (chRCC)*

Chromophobe is a less aggressive from of cancer compared to ccRCC, accounting for around 3-5% of renal epithelial cases with a mortality rate of less than 10% (Lopez-Beltran et al., 2006; Faisal et al., 1997). chRCC is also associated with the development of sarcomatoid features and metastasis in around 10% of cases

(Stec et al., 2009; Yusenko et al., 2009), while the autosomal dominant Birt Hogg Dubé syndrome predisposes sufferers to chRCC and oncocytoma (Hagenkord et al., 2011)

Pathologically, chromophobe presents as a solid orange tumour derived from the intercalated cell of the collecting duct with a mean size of 6 cm. Histologically, chromophobe cells are slightly opaque, and present with either pale cytoplasm in the majority of cells or eosinophilic cytoplasm (Stec et al., 2009). It has very similar appearance and histology to the benign renal neoplasm oncocytoma making differentiation of the two subgroups challenging (Pires-Luís et al., 2015). Due to the malignancy of chRCC accurate early diagnosis is vital to the appropriate course of treatment.

Immunohistochemistry has identified some variations between oncocytoma and chromophobe histology, such as, CK7, which stains strongly in chRCC but weakly in oncocytoma or kidney-specific cadherin, which is almost exclusively expressed in chRCC. However the immunohistochemical method of diagnosis and differentiation of the two histologies is not considered to be reliable or accurate (Ng et al., 2014). Chromosomal loss is a common feature of chRCC with the loss of between 5-13 chromosomes being reported in 75-100% samples. Chromosomal monosomy of chromosomes 1, 2, 6, 10, 13, 17, and 21 is the main feature in the majority of chRCC samples 58-93% of cases, while slightly less frequent are losses in chromosomes 3, 5, 8, 9, 11, 18 (12%–40%)(Cairns, 2010; Davis et al., 2014; Faisal et al., 1997; Yusenko et al., 2009).

Karyotyping of tumours is one potential method to distinguish between chRCC and oncocytomas, as chRCCs are often found to harbour loss of 2 or more

chromosomes (predominantly chromosomes 2, 10, 13 and 17 in 93%, 93%, 87% and 70% of chRCC, respectively,); these chromosomes are less frequently lost in oncocytomas, hypothetically enabling a distinction to be made between the two forms. However, karyotyping is regarded to lack the reliability and specificity required to make concise distinction between the two subtypes. (Ng et al., 2014; Yusenko et al., 2009).

Although the most common RCC susceptibility gene, *VHL* is not known to be mutated in chRCC (Cairns, 2010), other studies have identified significant number of mutations in the *TP53, PTEN, FAAH2, PDHB, PDXDC1* and *ZNF765* genes, with further recurrent mutations (not statistically significant) in; *ARID1A* and *PRKAG2*. Mutations in *PRKAG2* are suggested to result in gain of protein function, causing activation of *AMPK*, ultimately leading to the activation of an array of downstream pro-tumourigenic signals, including activating cellular energy homeostasis, regulating cell growth and metabolism and regulating cell polarity (Durinck et al., 2015).

Information provided on the cancer genome atlas database (TCGA) showed that investigations into the molecular classification of chRCC undertaken by Davis et al. on the 66 chRCC showed frequent mutations in *TP53* (32% cases) and *PTEN* (9% cases) with additional mutations reported in *mTOR* (3%), *NRAS* (1.5%), and *TSC1* or *TSC2* (4%) (Davis et al., 2014). However, it is not known whether the mutations identified in the tumours studied were actively pathogenic, or whether they had simply been acquired as passenger mutations. These findings concurred with a study by Durinck et al 2015, who investigated the genomic differences between pRCC chRCC and oncocytoma. In their study they also

reported finding frequent mutations in *TP53* (21.3% of samples) and *PTEN* (6.4% of samples) (Durinck et al., 2015).

Other observations of note made by Davis et al included frequent mutations in mitochondrial DNA particularly associated with the electron transport chain (18% samples) and the identification of a known cancer associated SNP in the *TERT* promoter in 4.5% of chRCC sample and abnormal structural variation of the *TERT* promoter in 12% of samples (see section 5.1.2) (Davis et al., 2014).

1.4.2.1.4 *Renal oncocytoma*

Oncocytoma is a benign and slow growing neoplasm of the kidney and accounts for around 5% of all renal neoplasms. Similarly to chRCC, it is predicted that oncocytomas are derived from cells of the intercalated cells of the distal tubule or collecting duct within the kidney. The tumours are usually large and often appear to be mahogany/brown or pale yellow in colour with a central scar (Faisal et al., 1997; Lopez-Beltran et al., 2006). Histological analyses have identified the cancer cells to be large with granular eosinophilic cytoplasm densely populated with mitochondria (Lopez-Beltran et al., 2006; Ng et al., 2014; Yusenko et al., 2009). Although these tumours are histologically similar to chromophobe, renal mass biopsy and imaging currently is the most effective method to diagnose oncocytoma prior to surgery (Ng et al., 2014).

Little investigation has been conducted upon the genetics of oncocytoma. Of the studies that have been published, it has been identified that the most prevalent chromosomal abnormalities are often loss of chromosome 1 (33% of samples) or chromosome Y (29% of samples), however, there have also been reports of translocations at 11q13 (Faisal et al., 1997; Yusenko et al., 2009). Recurrent

mutations are thought to be uncommon in oncocytomas, however NGS studies have recently identified mutations in *ERCC2* and *C2CD4C* to be associated with oncocytoma; although the biological role that these mutations play in tumourigenesis is yet to be ascertained (Durinck et al., 2015). Other studies investigating the increased density of mitochondria observed in oncocytomas have identified frequent mutations within the mitochondrial DNA of sporadic oncocytomas. It was therefore hypothesised that these mutations result in defective mitochondria, altering the cells energy production homeostasis and metabolism (Joshi et al., 2015; Lang et al., 2015).

Most investigations involving oncocytomas have focused on identifying key molecular and genetic signatures to help differentiate between chRCC and oncocytoma. For example; investigations into the microRNA (miRNA) profile of the two cancers have identified statistically significantly and differentially expressed miRNAs, and these can apparently be used to differentiate between oncocytomas and chRCC with relatively high efficiency. This included comparisons of the expression levels of 12 miRNA, analysed computationally to calculate the tumour subtype (Youssef et al., 2011).

Further investigations into differentiating the chRCC histology from that of oncocytoma have focused on the expression profiles of the two subtypes, which have revealed, differentially expressed genes; in particular *ADAP1, SDCBP2, HOOK2, BAIAP3* and *SPINT1* have been found to be up-regulated in chRCC while *ITGB3, MINOS1-NBL1* and *ASB1* appeared to be upregulated in oncocytomas (Durinck et al., 2015). Although these findings represent progress in the attempt

to clearly and reliably distinguish between the two tumours, as of yet, a robust and reliable method to achieve this has yet not been established.

1.4.2.1.5 *Sarcomatoid renal cell carcinoma (sRCC)*

It should be noted that sarcomatoid RCC is not a distinct RCC subtype, rather a feature of advanced tumour progression. However, due to the inclusion of clear cell derived sarcomatoid RCC samples in chapter 4 of this thesis, the current understanding of the sRCC histology and genetics shall be introduced below.

Sarcomatoid renal cell carcinomas (sRCC) are large tumours (≈10cm), characterised pathologically by a cell type similar to that observed in sarcoma (densely populated elongated, spindle-shaped cells, with a typical cellular morphology), and other characteristics associated with aggressive tumours such as necrosis and micro-vascular invasion(Shuch et al., 2012a). Jones et al., 2005 identified through analysis and observation that similar patterns of allelic loss and X-Chromosome inactivation patterns were present in sRCC and the surrounding ccRCC, which led to the conclusion that sRCC is derived from the pre-existing RCC. As a result, the majority of sRCC tumours are heterogeneous, with the parent tumour cell type most frequently observed.

sRCCs were originally classified as a distinct tumour type, however subsequent investigations have found sRCC to be derived from many RCC histologies. This finding then altered the classification of sRCC from a distinct tumour type to a tumour characteristic that can be used as an indicator of cancer progression (Faisal et al., 1997; Lopez-Beltran et al., 2006).

The presence of sRCC within a renal cell carcinoma is associated with advanced late stage tumours and therefore metastasis is a common observation (Delahunt

et al., 2013; Yan et al., 2015). It is predicted that sRCC is present in approximately 16-20% advanced renal cancers (Shuch et al., 2012a; Yan et al., 2015). Clinically, over 90% sRCC patients present with symptoms including with pain and haematuria (Humphrey, 2012). Moreover the presence of sRCC is associated with extremely poor prognosis; patients diagnosed with sRCC have a mean survival of between 4-9 months after diagnosis, and only a 15-22% five-year survival rate (Delahunt et al., 2013). Studies into survival of patients with sRCC identified that a high percentage of sRCC features is associated with poor prognosis and survival (Adibi et al., 2015; Mian et al., 2002). Furthermore it has also been reported that underlying histology from which the sRCC is derived has an implication on survival, with ccRCC derived sRCC conferring a lower risk of cancer-specific cell death compared to the non-clear cell subtypes (Nguyen et al., 2015).

The degree of sarcomatoid histology within a RCC tumour can vary greatly from as low as 1% up to 100% of the tumour histology. The percentage of sarcomatoid differentiation is usually determined visually by a trained pathologist. As mentioned previously, a negative correlation has been reported between the degree of sarcomatoid differentiation and prognosis (Kim et al., 2015).

Samples where the parent tumour can not be ascertained or sarcomatoid tissue forms 100% are defined as unclassified RCC however, these samples only account for between 2%–10% of all sRCCs (Humphrey, 2012; Shuch et al., 2012a). Most commonly sRCC is derived form ccRCC, this form of sRCC accounts for >80% of all sRCC identified (Shuch et al., 2012a) however only between 1 - 8% of all ccRCC tumours diagnosis are identified to have a positive sRCC histology (Humphrey, 2012). For other RCC histologies from which the

sarcomatoid histology can derive, sRCC features have been observed in 25-29% of all collecting duct RCCs, 2 – 9% of all chRCCs and 2 – 5% of all pRCCs (Delahunt et al., 2013).

Although limited, some molecular analysis has been undertaken on sRCC. Frequent loss of chromosomes 13q (75% of sRCC cases) and 4q (50% of sRCC cases) has been reported while chromosomal gains were observed with chromosome 17 (33% of cases) and chromosome 8q (25% of sRCC cases) (Humphrey, 2012; Jiang et al., 1998).

 Some mutational studies have reported frequent mutations in the *TP53* gene in up to 78.5% of sRCC samples, which were not observed within the neighbouring ccRCC tissues. With 72% of the mutated sRCC presenting with a 'hotspot' mutation in exon 8 codon 278 (CCT-CTT) (Oda et al., 1995), suggesting that these mutations are specific and perhaps driving the formation of the sarcomatoid histology. However immunohistochemical analysis of P53 expression on sRCC cells has failed to identify alterations in p53 immuno-reactivity between sRCC and the parent cancer (in this study either ccRCC or pRCC) which would be expected if there was a loss of protein function. This suggests that p53 may not be a key feature in sRCC differentiation (Kanamaru et al., 1999). More recent mutational studies have not directly focused on sRCC positive cancers specifically as a subgroup, however they have made observations about the histology within the study. For example it was observed that mutations within *JARID1c* were significantly associated with sRCC positive ccRCC (Freier et al., 2010). Another study investigating the prognosis and survival of ccRCC patients with *BAP1* mutations, identified that tumours possessing *BAP1* mutations were

also more likely to possess sarcomatoid histology and other aggressive features (Kapur et al., 2013). Although not directly analysing sRCC positive ccRCC samples, these studies suggest that there may be a genetic constituent to sRCC progression.

Presently diagnosis of sarcomatoid histology is determined by core biopsies of the tumour followed by visual analysis for sarcomatoid-like cells. However, limitations of biopsies, including small sample size which may exclude phenotypically relevant cells when sampling a large tumour and tumour heterogeneity may impede accurate diagnosis, calculation of sarcomatoid prevalence and ultimately prognosis predictions (Shuch et al., 2012a)

1.4.2.2 *Familial RCCs*

Familial RCC, also known as hereditary RCC is estimated to account for only 3% of all RCCs and unlike sporadic RCC, is usually characterised by an earlier age of onset and diagnosis and bilateral and/or multifocal tumour formation (Maher, 2011). Several inherited syndromes have been associated with a predisposition to familial RCC. Investigating these syndromes and the genes and pathways involved in them can provide invaluable insights into the genetic and metabolic basis of both familial and sporadic RCC. Some examples of well characterised familial RCC syndromes are summarised below (Johansson et al., 2013; Linehan et al., 2009).

1.4.2.2.1 *Von Hippel-Lindau Disease (VHL)*

VHL is the most common inherited syndrome associated with familial RCC, in particular ccRCC. VHL is an autosomal dominant familial neoplastic disease, with a prevalence of 1:35,000 live births. Individuals affected with VHL disease have a

>70% chance of developing ccRCC within their lifetime (Maher, 2011). It is reported that VHL syndromes are due to the inheritance of a pathogenic mutation in 80% of cases, while the remaining 20% are due to de-novo mutations (Hagenkord et al., 2011; Maher, 2013). Clinically characterised by systemic tumour formation, features frequently observed in VHL disease, include; retinal and cerebellar haemangioblastomas, phaeochromocytomas and pancreatic tumours and epididymal cysts (Gossage et al., 2015; Johansson et al., 2013; Maher, 2011). The genetic basis of VHL disease has been shown to be associated with disruptions to the tumour suppressor gene *VHL* (Latif et al., 1993)*,* either through germline mutations or through somatic inactivation (e.g. by LOH or DNA methylation) frequently reported in virtually all (≈ 100%) VHL families (Johansson et al., 2013; Linehan et al., 2010).

1.4.2.2.2 *Hereditary papillary RCC*

Hereditary pRCC is a rare autosomal dominant syndrome with an incidence of approximately 1 in every 10,000 people. (Maher, 2011). Classified according to the underlying genetics and histology observed. For example, hereditary papillary renal cancer type 1 (HPRC) is associated with pRCC type 1 histology, while hereditary leiomyoma renal cell carcinoma (HLRCC) is associated with pRCC type 2 histology (Section 1.4.2.1.2) (Johansson et al., 2013). HPRC is associated with an inherited alterations affecting the proto-oncogene *MET,* which lead to protein overexpression through mechanisms such as gain of function mutations, epigenetic deregulation or gene amplification (Courthod et al., 2015; Johansson et al., 2013; Maher, 2011). Reports have indicated that tumours are isolated to the kidneys only (Hagenkord et al., 2011). HLRCC on the other hand is associated with mutations in the gene *FH* at 1q42-43 (Courthod et

al., 2015; Linehan et al., 2010). *FH* encodes the *fumarate hydratase* enzyme, which is involved in the citric acid (TCA) cycle and protecting the cell against oxidative stress (Ooi et al., 2013).

Clinically HLRCC is characterised by multiple cutaneous leiomyomas and uterine leiomyomas (fibroids) at a young age in up to 90% of affected females (Johansson et al., 2013). Although HLRCC does predispose to pRCC, the formation of renal tumours is relatively uncommon. However, HLRCC kidney tumours are an aggressive and volatile tumour type, known to have the ability to quickly metastasize; for this reason, immediate surgical removal is usually advised (Johansson et al., 2013; Maher, 2011).

1.4.2.2.3 *Birt-Hogg Dubé (BHD) syndrome*

BHD is another autosomal dominant syndrome, characterised by benign skin tumours (fibrofolliculomas) on the face and upper body. Over 80% of patients affected with BHD also harbour lung cysts and have an increased risk of spontaneous pneumothorax, in addition to a predisposition to develop bilateral renal tumours (Murakami et al., 2007). BHD occurs as a result of germline mutations in the *FLCN* gene. The majority of patients harbour mutations within a hyper-mutable (hotspot) region of the *FLCN* gene in exon 11 (Johansson et al., 2013; Murakami et al., 2007; Nickerson et al., 2002). BHD associated RCC histology is typically reported to be made up of chRCC (33% of cases), oncocytoma (5% of cases) or more frequently a mix of the two histologies (chRCC/oncocytoma mix 50% of cases) (Hagenkord et al., 2011). However, histopathology of BHD RCC is diverse with ccRCC and pRCC subtypes also reported (Courthod et al., 2015; Maher, 2011).

1.4.2.2.4 *Constitutional chromosome 3 translocations*

Familial RCC associated with constitutional chromosome 3 translocations occur in a rare autosomal dominant condition associated nonsyndromic ccRCC. This condition is diagnosed via karyotyping. The first reported chromosome 3 translocation linked to familial ccRCC was t(3; 8)(p14;q24), although since this discovery, eleven different constitutional chromosome 3 translocations associated with ccRCC susceptibility have been identified (Maher, 2011). Investigations into chromosomal breakpoints have resulted in the further identification of numerous candidate tumour suppressor genes, including *Nore1* (alias *RASSF5*), *LSAMP* (Chen et al., 2003) and *FHIT* (Kvasha et al., 2008). All of which have also been reported to be frequently methylated in RCC.

However not all translocation breakpoints span a candidate TSG, so there are likely to be alternative mechanisms such as the chromosomal instability, which is the driving factor behind constitutional chromosome 3 translocations RCC formation (Maher, 2011).

1.5 Epigenetics of RCC

Historically, investigations into the genetic and epigenetic profiles of RCC have predominantly been undertaken by a candidate gene approach focusing on ccRCC or familial RCC. Subsequent investigations undertaking either a candidate and/or global approach to analysis, have identified numerous hypermethylated genes and epigenetic modifications associated with RCC and RCC subtypes. Providing an insight into both the genetic and epigenetic profiles of RCC subtypes (Henrique et al., 2012; Maher, 2013; Morris and Maher, 2010). An example of the most noteworthy or frequently reported genes methylated in RCC, include the TSG *VHL* (Latif et al., 1993). Which despite frequent mutations

and LOH in sporadic ccRCC (see section 1.4.2.1.1), is also frequently observed to be inactivated by DNA methylation of the promoter region in approximately 19% of ccRCC samples (Dulaimi et al., 2004; Herman et al., 1994; Kluzek et al., 2015). However, frequent methylation of *VHL* is not observed within pRCC, chRCC and oncocytoma (Dulaimi et al., 2004). However, methylation of other genes associated with inherited RCCs (*FH*, *FLCN*, *MET* but excluding *VHL*) are not frequently methylated (Maher, 2013).

Another example of a frequent methylated TSG is RCC is *RASSF1A*, located on chromosome 3p21.3, with regulatory roles in cell growth regulation via pro-apoptotic signalling, cell cycle arrest and the Hippo pathway signalling (Lopez-Beltran et al., 2010; Richter et al., 2009). *RASSF1A* has been identified to be frequently methylated in many RCC subtypes including ccRCC (46 - 60% samples) (Costa et al., 2007; Dulaimi et al., 2004; Maher, 2013; McRonald et al., 2009; Morris and Maher, 2010; Morrissey et al., 2001), chRCC (17% of samples) and oncocytoma (14% of samples)(Dulaimi et al., 2004). However, the highest frequency of *RASSF1A* promoter methylation was associated with pRCC (70-100% samples) (Costa et al., 2007; Dulaimi et al., 2004; Ellinger et al., 2011a; McRonald et al., 2009; Morris et al., 2010), but mutations within *RASSF1A* are rare in all RCC subtypes (Morris et al., 2011). Additionally other members of the *RASSF* family including *RASSF5 (Nore1)* have also been identified to be frequently methylated in 32% of sporadic ccRCC (Chen et al., 2003; Morris and Maher, 2010), suggesting that loss of this family of genes may have a role in tumourigenesis.

Further examples of frequently epigenetic silenced genes through DNA methylation in RCCs, include other of other chromosome 3 TSGs (such as *TU3A,*

DLEC1 and *FHIT*) (Maher, 2013), WNT pathway inhibitors (particularly members of the *DKK* and *SFRP* families of WNT inhibitors) (Maher, 2013; Morris et al., 2010) and MET inhibitor *SPINT2,* methylated in 30% and 40% of ccRCC and pRCC promoter regions respectively (Morris et al., 2005) . These reports support current understanding that DNA methylation plays a dynamic and complex role in RCC tumourigenesis (Heyn and Esteller, 2012; Maher, 2013).

The advent of high throughput technologies including microarrays, expression arrays, methylation arrays and next generation sequencing coupled with multi-centre collaborative projects such as The Cancer Genome Atlas (TCGA) has permitted large scale genome-wide analysis, profiling and investigation of large cohorts of RCC subtypes (ccRCC, pRCC and chRCC). Linking together aberrant genetic mutation, expression and methylation profiles with the aim of gaining a clearer understanding of a tumours biology and the genomic factors driving tumourigenesis (Davis et al., 2014; TCGA, 2016, 2013). Approaches such as this have permitted the identification of recurrent mutations in genes encoding chromatin regulatory proteins such *PBRM1, SETD2* and *BAP1* in ccRCC; where mutations have been associated with remodelling of the tumour cells epigenetic profile and altering the gene expression profiles of the tumour (Kluzek et al., 2015; TCGA, 2013). However, as previously mentioned no gene silencing DNA methylation has been observed for the aforementioned genes in ccRCC (Ibragimova et al., 2013a; Lasseigne et al., 2014).

Investigations have also been conducted to identify the relationship between histone modifications and RCC. Mutations within *SETD2* have been associated with a global decrease in histone modification H3K36me3 (Kluzek et al., 2015), a histone modification associated with active transcription and maintenance of

chromatin structure (Sims III and Reinberg, 2009; TCGA, 2013) . Studies attempting to identify prognostic histone markers reported poor prognosis in metastatic RCC patient presenting with high levels of H3K9me2 (associated with gene repression), (Harb-De la Rosa et al., 2015; Ramakrishnan and Pili, 2013), while low levels of H3K4 me1, 2 and 3 (associated with transcriptional activation) were associated with high grade, metastasis and poor survival (Ellinger et al., 2010). Additionally the study by Ellinger et al, compared the global H3K4 methylation state between RCC histologies including (ccRCC, pRCC, chRCC, sRCC and oncocytoma), in which they observed similar levels of H3K4me1 and 2 for ccRCC, pRCC, chRCC and sRCC, but higher H3K4me1 and 2 levels for the benign oncocytoma (Ellinger et al., 2010). Following the identification of recurrent mutations in histone methyltransferases *SETD2* in ccRCC and differential histone methylation patterns between oncocytoma and other RCCs, a recent study has explored the differential expression of histone methyltransferases as a method for distinguishing oncocytomas from chromophobe. Here they propose that increased expression of histone methyltransferases *SMYD2* in chRCC can distinguish chRCC from oncocytoma with a sensitivity of 71.0% and specificity of 73.3%, thus potentially being of use as a novel biomarker (Pires-Luís et al., 2015).

Detection of abnormal DNA methylation has been observed in RCC patients plasma and urine samples. Therefore it may be proposed that successful characterisation of the epigenetic and DNA methylation profiles of the different RCC subtypes within urine or plasma samples may provide a non-invasive method for diagnosis (Maher, 2013). Although many studies have attempted to characterise and define the different epigenetic profiles of RCC subtypes (Costa

et al., 2007; Dulaimi et al., 2004; Ellinger et al., 2010; Kluzek et al., 2015; Maher, 2013; McRonald et al., 2009; Morris et al., 2011, 2005; Ricketts et al., 2013; Youssef et al., 2011), or differentiate RCC tumours from normal kidney tissue via the DNA methylation profile (Lasseigne et al., 2014), as yet no definitive epigenetic profile, which can distinguish one RCC subtype from another, has been agreed upon. As yet no clinical application of epigenetic profiling has been approved for the differentiation RCC subtypes, and such an application currently remains far off (Kluzek et al., 2015).

1.6 <u>Treatment of RCC</u>

Traditional chemotherapy and cytotoxic drugs is of limited success in the treatment of RCC (Bylow et al., 2009; Stec et al., 2009). Currently surgical removal via partial or full nephrectomy is widely considered the gold standard and most effective course of action (Jonasch et al., 2014). For RCC patients in the UK, the National Institute for Health and Care Excellence (NICE), recommend the full (IPG136) or partial (IPG151) nephrectomy or ablation techniques (e.g. Laparoscopic cryotherapy (IPG405) and radiofrequency ablation (IPG353) as treatment for small and low grade RCC, and targeted treatment with pazopanib a tyrosine kinase inhibitor (TKI) for advanced RCC (TA215) ("Renal cancer | Guidance and guideline topic | NICE," 2016).

Targeted therapy has shown some success in improving patient outcomes in advanced RCC, compared to treatment with cytotoxic drugs and immunotherapy treatments such as interleukin-2 (IL-2) (Escudier, 2012; Shuch et al., 2015). As a result, targeted therapy is now the first pharmaceutical line of treatment for advanced RCC with several therapies including the anti-VEGF molecules

(sunitinib Tyrosine-kinase inhibitor (TKI), pazopanib (TKI) and bevacizumab (monoclonal antibody) which inhibit angiogenesis and an mTOR inhibitor (temsirolimus) having been approved for clinical use (Garajová et al., 2015; Linehan and Ricketts, 2014; Shuch et al., 2015). However, for many patients, treatment with these therapies fails to produce prolonged clinical responses. Current investigations to combat this resistance to treatment are examining the benefits and limitations of combination therapy. The use of MET and VEGF targeted inhibitors when used in combination has shown promising results in treating RCC, however the adverse effects are pronounced (Garajová et al., 2015). Other emerging therapies include the use of monoclonal antibodies, particularly monoclonal antibodies specific for immune check-point blockade molecules such as CTLA-4 (cytotoxic T lymphocyte antigen 4) and PD-1 (programmed cell death 1). These are currently in clinical trial and displaying promising clinical activity, however there is also high toxicity associated with these treatment (Jonasch et al., 2014 and references within)

The above treatment options are approved for all malignant RCCs, however the different RCC subtypes, may effect the therapeutic response. For example pRCC has been shown to be particularly resistant (more so then ccRCC) to chemotherapy and cytotoxic drugs (Bylow et al., 2009), furthermore the TKIs and mTOR inhibitors approved for malignant RCC have been reported little effect on the pRCC (Shuch et al., 2015). For chRCC, few clinical trials have been conducted evaluating therapy response in patients (Stec et al., 2009), although mTOR inhibitors have reported to have some clinical success (Shuch et al., 2012b) . However a surgical removal is still regarded as the best treatment and is associated with a 83% recurrence free survival rate after 5 years (Vera-Badillo et

al., 2012). With regard to other RCC subtypes such as sRCC, it has been suggested that cytotoxic chemotherapy maybe more effective for tumours with a high sarcomatoid percentage, and targeted therapy is more beneficial for tumours with a lower sarcomatoid percentage (Haas et al., 2012; Shuch et al., 2015). A phase 2 clinical trial combining targeted anti-angiogenic therapy (sunitinib) in combination with cytotoxic chemotherapy for sRCC patients exhibiting a high sarcomatoid content, has shown positive clinical activity, and is generally well tolerated with limited toxic effects (Michaelson et al., 2015).

Finally for oncocytoma patients, following diagnosis the course of treatment tends to be one of surveillance, with surgical removal performed on patients who's tumours have become too large and symptomatic (Neuzillet et al., 2005).

It is therefore important both economically and for the patients' prognosis and well being that the correct subtype of RCC is diagnosed quickly and effectively. So the most efficient and appropriate cause of treatment can be initiated, as always with, cancer, the earlier the intervention the better the prognosis.

1.7 Aims of the research

At the onset of this study in 2011, investigation and comparison of the epigenetic or genetic profiles or variations of the rare renal cancer subtypes, chromophobe RCC, renal oncocytoma and clear cell derived sarcomatoid RCC (sRCC) was limited. To the best of our knowledge no investigation into the DNA methylation state of clear cell derived sRCC had been conducted, a recent PubMed search identified that most sRCC based papers, are reports on histological observations and case studies [NCBI PubMed search, Keywords; Sarcomatoid Renal Cell Carcinoma sRCC; conducted on 14th August 2015].

For chRCC and oncocytoma some DNA methylation studies have been conducted on candidate genes, comparing the distinct histologies, however the small number of candidate genes studied did not yield any differential methylation profile between the two histologies (Costa et al., 2007; Dulaimi et al., 2004). Another study attempting to address the global methylation profile of chRCC and renal oncocytomas using BACS clones, identified similar DNA methylation profiles for the two histologies, with small regions of DNA methylation alterations that are specific to each neoplasm. However, further information as to the genomic locations and genes epigenetic regulation was lacking (Arai et al., 2011). Other studies which directly compared chRCC and oncocytoma focused on comparison of DNA copy number profiles (Yusenko et al., 2009), single nucleotide polymorphisms (Tan et al., 2010), gene expression (Pires-Luís et al., 2015; Tan et al., 2010), miroRNA profile (Fridman et al., 2010; Youssef et al., 2011) and immunohistochemical differences of the two histologies (Ng et al., 2014). These studies provided further evidence that chRCC and oncocytomas are distinct RCC subtypes. However, understanding of the two RCC subtypes global epigenetic and genetic profiles is still inadequate and until further understood classification and differentiation of the two subtypes by genetics profiling remains an aspiration. Therefore, the overall focus of this thesis was to address this gap in knowledge, to investigate and characterise the genetic and epigenetic profiles of these rare RCC histologies (Research published in the subsequent four years is further discussed in relevant thesis chapters). The hope was to identify features, which may aid diagnosis of specific subtypes permitting the development of RCC subtype, tailored treatment plans and eliminating the current methods of pathological, cytogenetic and immunochemical analysis

which are widely regarded as lacking reliability and specificity (Ng et al., 2014; Yusenko, 2010). Furthermore, identification of genetic variation may enhance understanding of the deregulated pathways promoting specific subtype tumourigenesis.

The main aim of the thesis was subdivided into the following three-pronged format, each chapter will introduce the unique techniques utilised within that chapter and address the following objectives;

1) To investigate and characterise the global DNA methylation patterns of chRCC and renal oncocytoma and to identify differential patterns of methylation between chromophobe and oncocytomas using the high throughput Illumina Infinium HumanMethylation450 BeadChip array and to compare our findings with the methylation data available from the TCGA.

2) Undertake a targeted approach to characterise the methylation profile of the *RASSF* family of TSG in the advanced and aggressive RCC histology, sRCC. With further investigation to ascertain if there is any association between methylation status and clinical data.

3) Identify the mutation profile of renal oncocytomas using whole exome sequencing and bioinformatic analysis, with the hope of identifying recurrent somatic mutations that may drive tumour formation. By the time this study was conducted the TCGA had published a multiplatform profile of chRCC including WGS (Davis et al., 2014), therefore the mutational landscape of chRCC was not studied in this thesis.

<div align="center">Chapter Two: **Methods and Materials**</div>

2.1 Samples

2.1.1 Renal cell lines

The kidney tumour cell lines used in this study were kindly provided by Professor Eamonn Maher and Dr Michael Lerman. RCC4 cell line was a kind gift from CH Buys. The origin and the establishment of the cell lines used in this study are described in the following references: Anglard et al., 1992; Ebert et al., 1990; Grossman et al., 1985. HEK293 cells were purchased from American Type Culture Collection (ATCC) (ATCC Number: CRL-3216). Mr Dean Gentle generously extracted DNA from 18 ccRCC cell lines (RCC1, RCC4, RCC11, RCC12, UMRC2, UMRC3, KTCL26, KTCL140, Caki 2, SKRC18, SKRC39, SKRC45, SKRC47, SKRC48, SKRC54, A498, A704, 796P) (section 2.6.5), and RNA extracted from 4 ccRCC cell lines (UMRC3, KTCL140, RCC11, SKRC45) (section 2.6.6) with and without 5-Aza-2'-Deoxycytidine treatment (a DNA methyltransferase inhibitor which results in DNA demethylation) (section 2.6.4).

2.1.2 Primary DNA samples

Normal age matched kidney DNA was previously extracted from control kidney samples retrieved during non-cancer related surgeries upon the kidney (mean age 57 years, range from 23-79 years).

Anonymised primary DNA from tumour and matched normal from the following histologies of renal tumours; oncocytoma, chromophobe, papillary clear cell and clear cell with sarcomatoid histology were kindly extracted and provided by our collaborators, Dr G Kovacs (Department of Laboratory Medicine; University of Pecs Medical School; Pecs, Hungary) who supplied the European cohort and Dr M Yao (Department of Urology; Yokohama City University School of Medicine;

<div align="center">58</div>

Yokohama, Japan) who provided samples for the Japanese cohort. The clinical data available for all primary samples is displayed in Table 2-1.

All DNA was stored at -20°C. Ethical approval was obtained from institutional and local research committees for the collection and use of all the samples. Research was conducted according to the principles expressed in the Declaration of Helsinki (World Medical Association, 1974).

2.1.3 DNA concentration

The concentration and purity of all DNA and RNA was determined via a Nanodrop spectrophotometer ND-1000 (Thermo Fischer Scientific Inc). The Nanodrop was primed and blanked with 2μl dH_2O prior to quantification of DNA samples. 2μl of the sample was loaded on the Nanodrop pedestal and quantified in ng/μl. Sample purity was determined by observation of 260/280 and 260/230 ratios; values of 1.6 – 2.0 and 1.8 - 2.2 were observed in our samples. Generally 260/280 and 260/230 ratios of 1.8 for DNA and 2.0 for RNA are accepted as "pure" for DNA and RNA samples respectively. Values appreciably lower than these indicate the presence of protein, phenol or other contaminants within the sample (Thermo Fischer Scientific Inc., 2008).

Histology	Gender	Average Age (Years)	Average Tumour Size (cm)	Sample	Ethnicity	UICC Stage (2002)	Metastasis	Average survival time (month)	Inclusion in project:		
									Chapter 3	Chapter 4	Chapter 5
Onco	F = 16 M = 12 n.d = 16	59.9 (30 - 76) n.d = 16	5.4 (1 - 14) n.d = 18	T/N = 34 T = 10	E = 32 J = 12	-	-	-	450k E = 18 450k J = 3	-	E Screen = 31 J Sceen = 4 J NGS = 5 E Mentioned = 1 J Mentioned = 3
chRCC	F = 22 M = 14 n.d = 6	53.4 (21 - 80) n.d = 6	7.0 (1.8 - 15.5) n.d = 6	T/N = 36 T = 6	E = 18 J = 24	I = 11 II = 7 III = 4 IV = 1 n.d = 19	n.d	n.d	450k E = 18 450k J = 2 CoBRA J = 22	-	-
ccRCC	F = 5 M = 15	61.3 (45 - 81)	6.3 (2.3 - 16)	T/N = 20	J = 20	I = 6 II = 3 III = 6 IV = 5	n.d	79.9 (6.6 - 156.3)	-	J = 20	-
sRCC	F = 9 M = 22	61.9 (33 - 78)	8.4 (2.5 - 20)	T/N = 25 T = 6	J = 31	I = 2 II = 1 III = 9 IV = 19	No Metastasis = 16 Matastasis = 15	42.1 (1 - 195)	-	J = 31	-

Table 2-1: Clinical data for primary RCC samples used in this study
Table summarises the clinical data for the primary tumours used within this study. Table details tumour: onco: oncocytoma, chRCC: chromophobe, ccRCC: clear cell RCC and sRCC: sarcomatoid RCC derived from clear cell RCC), the ethnicity of the samples (**J**: Japanese, **E**: European), patients age at diagnosis, gender of patient (**M** = male, **F** = female), tumour size and the chapter in which the samples were used (for chapter 3: 450K indicates the samples run on the methylation array, CoBRA indicated the samples used for secondary confirmation of methylation. For chapter 5, **NGS**: refers to the samples sent for next generation sequencing, **screen**: refers to the sample screened by Sanger sequencing, and **mentioned**: are samples that were screened by Sanger sequencing and explicitly mentioned within the chapter) . Whether the samples were tumours (**T**) only, or if matched normal (**N**) were available. If applicable, the grades of the tumour are displayed according to the 2002 UICC staging classification (stage I-IV) and whether metastasis was observed (see section 1.4.1), dashed line denotes data of this type not usually recorded. Patient survival time in months and the overall outcome of the patient is documented (where **cd** = cancer death, **awd** = alive with disease and **ned** = no evidence of disease). For all cases **n.d** = no data available.

2.2 <u>General techniques</u>

A key tool in the study of DNA is polymerase chain reaction (PCR), a technique employed to rapidly and accurately amplify regions of interest within a DNA sample for further analysis (Ishmael et al., 2008). Since its discovery, many different forms and adaptations of the traditional PCR have been developed, focussing on different areas of investigation. Below are detailed the key methodologies used throughout all the projects undertaken within the study.

2.2.1 Primer design

In order to employ the various PCR techniques in this study, primers were required to provide the starting template for the *Taq* polymerase (see section 2.3.2.1). The design of the primers varied depending on the PCR technique used, however all primers were designed to be specific for the region of interest and were usually between 18 and 30 nucleotides, with an average primer length of 20 nucleotides, see the following sections for more information. .

2.2.2 Methylation primers

Combined bisulfite restriction analysis (CoBRA) and Methylation Specific PCR (MSP) primers used in this study were designed manually unless otherwise specified. Primers were designed within the CpG islands as defined by University of California, Santa Cruz Genome Browser (genome.ucsc.edu). Preferably, the region interrogated was upstream of the transcription start site (TSS) and did not include the coding region. In the case of the work undertaken in Chapter 3, primers were designed to incorporate the CG dinucleotide of interest within the amplified region regardless of location to TSS or CpG island. All primers were

designed to be specific for bisulfite modified DNA. Dr Victoria Hill designed the CoBRA and MSP primers for the RASSF family of genes.

2.2.3 Combined bisulfite restriction analysis (CoBRA) Primers

CoBRA primers are designed to amplify a region between 300-500bp containing at least one DNA sequence motif specific for the digestions enzymes *BstU1* (5'-CGCG-3') or *Taqα1* (5'-TCGA-3'). This permits methylation specific digestion of the PCR product. To avoid primer bias, any C belonging to a CG dinucleotide within the forward primer region was labelled 'Y' (indicating during synthesis a 50:50 ratio of C:T at that particular nucleotide), for the reverse primer (the reverse complement) any G belonging to a CG dinucleotide was labelled 'R' (indicating during synthesis a 50:50 ratio of G:A at that particular nucleotide). To increase specificity the 3' region of the primer should incorporate a CAC or ACC, avoiding a CG dinucleotide at the 3' end. High proportions of lone C nucleotides increases the specificity of the primer for bisulfite modified DNA (Xiong and Laird, 1997).

2.2.4 Methylight primers

The Methylight primers and probes were previously designed by Dr Ashraf Dallol (Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Saudi Arabia) to amplify a small region of approximately 100 ± 20 bp. Methylight primers were designed to be specific to methylated DNA. Methylight primers require multiple CG dinucleotide within the primer sequence, particularly towards the 3' region to increase specificity and discriminate between methylated and unmethylated sequence (Davidović et al., 2014). To quantify methylation a 20bp probe was designed complimentary to the middle of

the amplified region, specificity for methylated DNA, with a fluorescent dye at the 5' end and a quencher at the 3' end.

2.2.5 Genomic and sequencing primers

Primers for exome sequencing and standard amplification of genomic DNA were designed using Primer 3 (Koressaar and Remm, 2007) and Primer Blast (Ye et al., 2012) to ensure specificity, prevent self-complimentary binding and primer dimers. PCR amplification regions were between 200-500 bp. Primers were between 18 and 25bp in length and contained approximately 50% GC content with a melting temperature (Tm) of between 58°C and 62°C with the forward and reverse primer preferably not exceeding a tm difference of 1 °C to increase reliability and specificity (Lorenz, 2012).

Primers were designed in the intronic region approximately 50 bp either side of an exon (boundary defined by Ensembl.org), or, in the case of specific mutations, primers were designed to include the base of interest. If the exon exceeded the maximum PCR amplification range, multiple primers were designed that overlapped to ensure all the exon was sufficiently investigated.

Single nucleotide polymorphisms (SNPs) present in genomic DNA can affect the accuracy of PCR amplification and generate a PCR bias, by altering efficiency of polymerase extension, or primer specificity. It was therefore imperative to ensure primers annealed to regions devoid of SNPs. To minimize the risk of SNP induced primer bias all primers were visually compared to known SNPs in Ensembl and checked using on line bioinformatics tool SNPcheck3 (Snpcheck.org)

2.2.6 Expression primers for RT-PCR

Primers specific for cDNA were designed to evaluate the expression of genes following RT-PCR (see section 2.7). These primers shall be termed expression primers. Expression primers were designed following the same rules as Genomic and Sequencing primers (section 2.2.5), using Primer Blast and Primer 3. However, in this case cDNA was used as the template and ideally at least one primer should span an exon:exon boundary.

2.2.7 Loss of heterozygosity (LOH) primers

To identify loss of heterozygosity (LOH), the nearest heterozygous micro-satellites either side of the gene of interest (according to UCSC) were identified. The West Midlands Regional NHS Genetics laboratory provided information on the probability of heterozygosity and primer sequences for each micro-satellite. Unlike conventional primers, a FAM fluorescent tag was added to the 5' region of the forward primer for detection.

2.2.8 PCR reagents and machines

All PCR reactions were completed on either Biometra thermo-cycler or Bio-rad Tetrad 2 Pelther Thermal cycler unless otherwise specified. Unless otherwise specified all reagents used in PCR reactions were supplied from Roche (Roche; Cat #: 12158264001) except for the dNTPs, which were from Thermofisher Molecular Biology (Thermofisher; IBR stores; Cat #: IBR10297018).

2.2.9 Agarose gels

Gel electrophoresis was employed to analyse PCR products. Electrophoresis separates DNA by fragment size, with the smallest fragments travelling through the gel matrix quicker, while the larger fragments travel slower and therefore a

smaller distance through the gel. The percentage of the agarose gel can be varied depending on the size of the DNA fragment, with a higher percentage being used to differentiate smaller fragments. Unless specified, all agarose gels used in this study were made at 2% (w/v).

By dissolving 2% (w/v) agarose powder (Bioline; Cat #: BIO-41025) in 200 ml TBE buffer (90mM Tris-base, 90mM boric acid and 2mM EDTA, pH 8.0) a 2% agarose gel was made. TBE buffer was diluted from 10x stock (National Diagnostics; Cat #: EC-860) with dH_2O prior to the addition of agarose gel. The agarose solution was heated until all powder had dissolved and solution was clear. Once cooled, but not set, 0.1µg/ml ethidium bromide (Sigma Aldrich; Cat #: E8751) was added. The agarose was then poured into a casting tray with comb and left to set at room temperature. Once set, the comb was removed leaving wells in which DNA could be loaded. Gels were placed in an electrophoresis chamber, submerged in TBE buffer, and DNA was loaded into wells. To aid the loading of DNA and to visualise the progress of the electrophoresis, 5µl loading dye was added to each DNA sample (Loading dye consisted of, 50% (v/v) glycerol (Sigma Aldrich; Cat #: G5516), 48.5% (v/v) distilled water, 1% (v/v) EDTA (Sigma Aldrich; Cat #: E9884 and 0.5% (w/v) Orange G (Sigma Aldrich; Cat #: O3756). In one well per row, 5µl 100bp DNA ladder (Invitrogen; Cat #: IBR15628019) was loaded as a size standard. Gels were then run at 170 V permitting the DNA to migrate through the gel towards the positive electrode, for as long as required to separate the PCR products. DNA bands were visualised under UV light and imaged using SynGene InGenius GelDoc system and Genesnap imaging software.

2.3 <u>Methylation studies</u>

Investigation of DNA methylation was undertaken on both an individual gene and a whole genome level basis. DNA methylation is not retained through PCR amplification, therefore all investigations undertaken in this project were performed on sodium bisulfite modified DNA.

2.3.1 Bisulfite modification

Bisulfite modification of DNA allows the methylation status of DNA to be determined by permanently altering the DNA sequence to reflect the methylation status of the DNA. The method of action of bisulfite modification is described in section 3.1.1.1. Bisulfite modification of DNA was performed using Qiagen EpiTect Kit (Qiagen; Cat #: 59104), following the manufacturer's instructions; all equipment and reagents were supplied in the kit unless otherwise specified. Aliquots of 'bisulfite mix' were completely dissolved in 800µl RNase-free water prior to setting up the reaction. Each reaction consisted of 1µg genomic DNA with 85µl of bisulphite mix and 35µl green DNA protection buffer (required to maintain and indicate pH during the reaction, blue colour indicates correct pH). Reactions were vortexed to ensure all reagents were homogenously distributed and run on a PCR machine with the reaction conditions shown in Table 2-2.

Step	Temperature (°C)	Time (Minutes)
Denaturation	95	5
Incubation	60	25
Denaturation	95	5
Incubation	60	85
Denaturation	95	5
Incubation	60	175
Hold	20	∞

Table 2-2: Bisulfite modification conditions

Post bisulfite conversion, the reactions were transferred to 1.5ml micro-centrifuge tubes (IBR stores; Cat #: IBR-EPP-1.5), and 560µl buffer BL and 10µg/ml carrier RNA was added to each sample. Samples were then briefly vortexed and centrifuged before transferring to an EpiTect spin column.

Samples were centrifuged at 16,300 x g for 1 minute at room temperature (all following centrifugation steps for bisulfite modification were performed with these conditions). Flow-through was discarded and column membranes were washed with 500µl buffer BW and centrifuged. Flow through was once again discarded. To de-sulphonate samples, 500µl buffer BD was added to each membrane and incubated for 15 minutes at room temperature, to avoid oxidation of the buffer it was imperative to ensure all lids were closed for the duration of this incubation. Following incubation, samples were again centrifuged and flow-through discarded and washed twice with 500µl buffer BW, between each wash, samples were again centrifuged and flow through discarded. EpiTect columns were then placed into fresh collection tubes and centrifuged to remove residual liquid. To ensure all liquid was removed from the membrane the EpiTect columns were transferred into fresh 1.5ml microcentrifuge tubes and incubated at 56 °C for 5 minutes. The bisulfite modified DNA was then eluted by the addition of 20µl buffer EB directly to the membrane and left to stand at room temperature for 5 minutes before centrifugation. To enhance the yield of bisulfite modified DNA obtained; the elution step was performed twice.

2.3.2 Candidate gene methylation analysis

To investigate the methylation status of candidate genes the following methods were conducted on bisulfite modified DNA.

2.3.2.1 *Combined Bisulfite Restriction Analysis (CoBRA)*

Methylation sensitive digestion enzymes such as *BstUI* (Fermentas UK; Cat #: *ER0921)* and *TaqαI* (New England Biosciences; Cat #: R0149S) were used to distinguish the methylation status of regions of interest. Digestion of the bisulfite modified DNA is only observed if methylation is present.

Each primary CoBRA reaction was conducted on 0.05μg bisulfite modified DNA. The PCR master mix consisted of 1x PCR buffer (diluted from 10x stock containing: 500mM Tris-HCL (pH 8.3), 100mM potassium chloride, 50mM ammonium sulphate, 2mM magnesium chloride), 2.5mM dNTP mix, 1x GC-RICH solution (from 5x stock), 0.8μM forward primer, 0.8μM reverse primer, and 0.5U FastStart *Taq* DNA polymerase (Roche; Cat #: 12158264001) made up to the final volume of 25μl with distilled water (dH$_2$O). Unless specified, a second round of PCR was conducted on 5μl of the PCR product from the primary CoBRA reaction in order to increase the yield. The secondary reaction was conducted in a 50μl volume using either semi-nested or fully-nested primers. The concentrations of PCR master mix components remained the same as for the primary reaction, with the only exception being the increase of FastStart Taq to 1U per reaction.

The PCR reactions were run using either a PCR protocol optimised for amplification of CoBRA PCR products dubbed here as 'methylation' PCR protocol (**Table 2-3**) or a touchdown PCR program (used to avoid amplification of

unspecific sequences; Table 2-4). Primer sequences, annealing temperature, PCR program, cycle numbers and digestion enzyme for each region investigated are listed in Table 7-1.

Following the CoBRA reactions, 5μl of PCR product was run by electrophoresis on an agarose gel, to determine a suitable volume of PCR product for digestion (between 5-20μl). To establish the methylation status, the predetermined volume of PCR product was digested either with BstUI (which recognises and cleaves DNA at CGCG) or Taqα1 (which recognises and cleaves DNA at TCGA) depending on restriction sites present.

BstUI digestion involved incubating the desired volume of PCR product with 1 U *BSTU1* and 1x Buffer R (10x stock; Fermentis UK supplied with *BstU1*). dH$_2$O was added to make the reaction up to a total volume of 25μl. Samples were incubated for 16 hours at 37 °C. Digestion via *Taqα1* involved incubation for 1 hour at 65°C of an appropriate amount of PCR product with 1U *Taqα1* per 5μl PCR product, 1x BSA (from 100x stock), 1x NE Buffer (from 10x stock) (all reagents supplied with the *Taqα1* enzyme) and dH$_2$O to make the total volume up to 25μl.   After digestion, samples were run on an agarose gel next to undigested PCR product.

Step	Temperature (°C)	Time (Minutes)	First round PCR	Second round PCR
Denaturation	95	5		
Denaturation	95	1		
Annealing	Ta	1	35 Cycles	40 Cycles
Elongation	72	2		
Final elongation	72	10		
Hold	10	∞		

Table 2-3: Methylation PCR conditions for primary and secondary PCR reactions

Step	Temperature (°C)	Time	First round PCR	Second round PCR
Denaturation	95	5 minutes		
Denaturation	95	45 seconds		
Annealing	Ta +5°C (Decrease 1°C per cycle)	45 seconds	5 Cycles	5 Cycles
Elongation	72	45 seconds		
Elongation	72	45 seconds		
Annealing	Ta	45 seconds	35 Cycles	40 Cycles
Elongation		45 seconds		
Final elongation	72	10 minutes		
Hold	10	∞		

Table 2-4: Touchdown PCR conditions for primary and secondary PCR reactions

2.3.2.2 *Methylight*

Methylight is a sensitive, quantitative PCR technique using real time PCR. Primers are designed to be specific for methylated sequences, while the incorporation of a methylation specific probe increases specificity and allows quantification of methylation.

A five point serial dilution was used to create a standard curve to quantify the results. This was done in duplicate, with fresh dilutions made for each Methylight run. A 1 in 5 serial dilution of EpiTect 100% *in-vitro* methylated control (Qiagen, Cat ♯: 59655) into nuclease-free water was created in order to achieve the final concentrations of 10ng, 2ng, 0.4ng, 0.08ng and0.016 ng per well. Bisulfite modified sample DNA was diluted to 5ng/µl with nucleases-free water before use. The Methylight PCR reaction was composed of 2µl (10ng)

70

bisulfite modified DNA or Standard curve dilution or 1µl (10ng) control DNA, 5µl TaqMan® Fast Universal PCR Master Mix (from 2x stock; Life technologies; Cat #: 4366072), 1µl probe/primer mix (from 10x stock consisting of 4µM forward and reverse primers and 2µM TaqMan® probe diluted in nuclease free water to achieve the final concentration of 0.4µM primers and 0.2µM probe) and 2µl nuclease free water, resulting in a total volume of 10µl. The reaction was run on Bio-Rad I cycler IQ5 multicolour PCR detection system (Table 2-5), where fluorescence was imaged and quantified during the amplification step, providing a Cycle Threshold (C_t) value for each sample.

Step	Temperature (˚C)	Time	
Denaturation	95	10 minutes	
Denaturation	95	15 seconds	
Annealing and image capture	60	1 minute	45 Cycles

Table 2-5: Methylight PCR conditions

Each Methylight reaction was conducted in a 96 well plate and included investigation of the gene of interest (GOI- FAM labelled) and a house-keeping gene as an internal control (Alu (VIC labelled)) on the same plate. As additional controls 10ng 100% *in-vitro* methylated positive control and 10ng EpiTect unmethylated control (10ng/µl; Qiagen; Cat #: 59665) were included for each primer set and each repeat. The five point standard dilution was only included for the house keeping reference gene.

To increase reliability and reduce margin of error, all samples, standards and controls were performed in duplicate on a 96 well plate and each Methylight reaction was repeated in triplicate. All primer sequences are described in Table 7-1.

2.3.2.2.1 *Calculating percent of methylated reference*

The percent of methylated reference (PMR) is a method of quantifying and rationalising the level of DNA methylation in relation to the 100% methylated control. The stages undertaken to calculate the PMR, were as follows; first the C_t values of duplicate samples were averaged. Then, as described by Dallol et al., 2011, a standard curve graph was produced by plotting the $\log_{10}$ standard concentration on the x axis against the C_t value recorded for each of the standards. The correlation of the standard curve was calculated via Pearson's correlation coefficient. An 'r' value of greater than 0.9 was considered of sufficient correlation to progress with the analysis of the samples. The correlation of the standard curve was assessed after every run in order to ensure reliability and accuracy of the run. In addition to calculating the correlation, the trend line of the standard curve was also calculated using linear regression (Equation 1);

$$y = mx + c$$

Equation 1: Equation for a straight line
Where *m* is the gradient and *c* is the y intercept.

The next stage was to calculate the quantity of each sample (Q) (Equation 2);

$$Q = e^{(\frac{ct-c}{m})}$$

Equation 2: Quantification of each Methylight sample
Where *c* and *m* are derived from the previously calculated equation for a straight line, and *e* is the mathematical constant 2.71828 the base of the natural logarithm (Equation 1).

Following the calculations of the above, the PMR was then calculated. This allowed the calculation of PMR by (Equation 3).

$$PMR = \left(\frac{\left(\frac{Q - GOI_s}{Q - Alu_s} \right)}{\left(\frac{Q - GOI_{cont}}{Q - Alu_{cont}} \right)} \right) \times 100$$

Equation 3: Percentage of methylated reference
Where:

GOI	= Gene of interest
Alu	=*Alu* house keeping gene
s	= sample
cont	=100% *in-vitro* methylated control
Q	= sample quantity

PMR values for each sample were averaged across the three repeat reactions. A sample was determined to be positive for methylation if the PMR was >10 (Buhmeida et al., 2011; Dallol et al., 2011).

2.3.3 Genome wide methylation array

Investigation of the global DNA methylation profile was performed using the Infinium HumanMethylation 450 BeadChip methylation array (450K array), a comprehensive technology that interrogates the DNA methylation profile of over 485,000 CpG loci, correlating to 99% coverage of RefSeq genes and 96% of CpG Islands (Illumina.com, 2015).

Analysis of genome wide methylation profile of 4 normal kidney, 21 primary oncocytomas and 20 primary chromophobe (1 oncocytoma sample sent in duplicate to assess the reliability and reproducibility of the assay) was undertaken using the aforementioned Infinium HumanMethylation 450 BeadChip methylation array. Samples were outsourced to J. Bauer, Cambridge

Genomic Services, Department of Pathology, University of Cambridge, UK to be run according to the manufactures instructions. In summary genomic DNA was bisulfite modified before hybridisation to the 450K array chip and interrogated following manufactures instructions.

Data were obtained and normalised using Illumina Genome Studio software. Due to the chemistry and technology employed by the 450K array, it was observed that data should be corrected to avoid colour bias. This was achieved by the modification of an existing quantile pipeline, formally developed for the Infinium HumanMethylation27 Bead-Chips (Du et al., 2008). In summary, the two probes utilised in the 450K array (Type I and II) (see section 3.1.2.1) were separated and normalised independently to omit any colour bias and to ensure both probe types possessed the same distribution on the array, prior to running on the standard normalisation pipeline contained within the Illumina Genome Studio software.

As part of the above pipeline the methylation status for each locus was then calculated and provided as an average β value (Equation 4), which compares the fluorescence intensity of methylation specific probes with unmethylated specific probes (Dedeurwaerder et al., 2011). A β value of 0 corresponds to completely unmethylated locus while a β value of 1 represents 100% methylation.

$$\beta = \frac{Intensity^M}{(Intensity^U + Intensity^M + 100)}$$

Equation 4: Calculation of the Infinium HumanMethylation 450 BeadChip β value
Where: *M*= Methylated and *U*= Unmethylated reading

When analysing the output data from the 450K array, a β value of greater than 0.5 was considered to represent methylation of that particular CpG locus, with cancer specific methylation classified as probes with a β value greater than 0.5 in tumour samples but less than 0.25 in all normal samples (Weisenberger et al., 2008).

2.3.4 Single colony sequencing

Single colony sequencing is a method used to amplify and isolate single alleles of DNA for Sanger sequencing to determine the DNA sequence of a region of interest. This is a particular useful method for analysing methylation of bisulfite modified DNA, as the level of methylation can differ between alleles.

2.3.4.1 *PCR amplification and gel extraction*

Genes and regions of interest were amplified on bisulfite modified DNA by CoBRA PCR (Section: 2.3.2.1). Following amplification, the full 50μl of the secondary PCR product was run on an agarose gel. Using a scalpel, each PCR product was excised, omitting as much agarose gel as possible without compromising the PCR band, and placed into individual labelled 1.5ml micro-centrifuge tubes (IBR stores; Cat #: IBR-EPP-1.5), ready for gel extraction via the QIAquick Gel extraction kit (Qiagen; Cat #: 28704). The gel extraction process involved weighing of the extracted agarose/PCR band, followed by the addition of 3 x mass (in mg) of the agarose fragment in volume (μl) of Buffer QG. Prior to incubation, tubes were vortexed regularly at 50 °C for 10 minutes, or until all the agarose had dissolved. Buffer QG contains a pH indicator and should remain yellow throughout the above step indicating a pH < 7.5. Once fully dissolved, a volume of 100% isopropanol (Sigma-Aldrich; Cat #: 190764) was added, the volume of isopropanol added in μl was equal to the mass of the agarose fragment

in mg. Samples were thoroughly vortexed prior to the solution being transferred to a QIAquick spin column, containing a membrane which binds the DNA, and a 2ml collection tube. Samples were then centrifuged at room temperature for 1 minute at 17,900 x g and flow-though was discarded. A further 500µl Buffer QG was added to each spin column and a second centrifugation step was conducted as before and flow-through was discarded. Samples were then washed by the addition of 750µl Buffer PE to the spin column, to ensure thorough washing and removal of salt, samples were left to stand at room temperature for 5 minutes before being centrifuged for a further minute using the same conditions as before. Flow-through was once again discarded. To remove any residual buffer, the centrifugation step was repeated. To elute the DNA, spin columns were transferred to a fresh 1.5 ml microcentrifuge tube, 30µl Elution Buffer EB was added directly to the column membrane and left at room temperature for 5 minutes to increase DNA yield. Samples were then centrifuged for a final time at room temperature for 1 minute at 17,900 x g to elute the purified DNA from the column. The purified PCR product was either used immediately or stored at -20 °C.

2.3.4.2 *Ligation*

Gel extracted DNA was ligated into the Promega pGEM-T Easy vector system (Promega; Cat #: A1360), a method utilised to introduce PCR products into bacterial cells for amplification by cloning. The ligation involved an overnight incubation at 4°C of 3.5µl purified gel extracted PCR product with 0.5µl pGEM vector, 1µl T4 DNA ligase, 1µl 10x ligation buffer and 4µl nuclease free water. The pGEM-T easy vector system was chosen due to the high ligation efficiency, ease of use and intrinsic selection processes once transformed into bacterial

cells. These selection processes included the presence of an ampicillin resistance gene, conferring resistance to the antibiotic, and a LacZ gene, which is disrupted upon successful insertion of the DNA fragment of interest.

2.3.4.3 *Preparation of LB agar plates*

Prior to transformation of the ligated vector into bacterial cells, LG agar plates were made to provide the nutrients and optimal conditions for the bacterial cell colony formation. Plates were prepared by mixing 14 g LB agar (Sigma Aldrich; Cat #: L2897) in 400 ml distilled water to obtain a concentration of 35 g/L. This solution was then autoclaved and allowed to cool sufficiently (but not set) before the addition of 100µg/ml ampicillin (Sigma-Aldrich; Cat #: H0166). Molten agar solution was then poured into petri dishes (approximately 20 ml agar/petri dish) and left to set. Once set, plates were stored upside down to minimise evaporation, at 4 °C until required.

2.3.4.4 *Transformation of bacterial cells and selection*

The ligated PCR/pGEM vector was transformed into α-select silver efficiency chemically competent cells (Bioline; Cat #: BIO-85026). Competent cells were defrosted on ice and aliquoted equally into 1.5ml micro-centrifuge tubes (to minimum volume of 50µl/ tube). 10% of the volume of the cells of PCR ligated pGEM vector was gently added (e.g. for 50µl of competent cells 5µl PCR-pGEM vector was added). Stirring gently mixed the samples, with care taken to minimise cell damage. The competent cells/PCR-pGEM vector mixture were then incubated on ice for 20 minutes, followed by a heat-shock step at 42 °C for 30 seconds, and a further 2 minute incubation on ice. 500µl SOC media (Invitrogen; Cat #: 15544034) was then added and samples incubated for 90 minutes at 37 °C with gentle agitation at 200 rpm on a shaking heat block.

During the above incubation period, LB agar (+Ampicillin) plates were covered with 80μl 0.1 M IPTG (Life technologies; Cat #: 15529-019) and 20μl 50 mg/ml X-gal (Bioline; Cat #: BIO37035) and permitted to dry. Transformed cultures were then plated onto the LB agar (+ Ampicillin, + IPTG, + X-gal) plates by gently adding 200μl of culture to the agar and distributing evenly around the plate using a plastic spreader (IBR stores; Cat #: IBR-GC543). Plates were then incubated upside-down at 37 °C for 16 hours.

The presence of the ampicillin within the agar permits only the growth of successfully transformed bacterial cells, as the pGEM vector containing the ampicillin resistant gene. Meanwhile the inclusion of X-gal and IPTG permits for selection of colonies containing the PCR insert within the vector, as successful insertion of the PCR product can be visualised by the formation of white colonies, due to the disrupted Lac-Z gene. Lac-Z encodes the β-galactosidase enzyme that hydrolyses X-gal and produces a blue pigment bi-product. When Lac-Z is disrupted the colonies remain white. The inclusion of IPTG assists in the transcription of the Lac operon, increasing the amount of β-galactosidase translated, resulting in enhanced hydrolysis of X-gal, emphasising the colour of the colonies.

White colonies were selected using a pipette tip. Colonies were transferred into 20μl dH$_2$0 and denatured at 95°C for 5 minutes.

2.3.4.5 *Single colony-PCR*

Following colony selection and denaturation, a single colony (SC) PCR was performed to confirm the presence of the PCR insert.

In summary, 7μl of denatured selected colony solution was PCR amplified in a 30μl reaction consisting of 1x PCR buffer, 2 mM magnesium chloride, 2.5 mM dNTP mix, 1x GC rich solution, 0.8μM forward primer (pGEM T4: TAATACGACTCACTATAGGG), 0.8μM reverse primer (pGEM-R: ACACTATAGAATACTCAAGC-3) and 0.5 υ FastStart Taq DNA polymerase and run on a PCR machine with the following specialist touchdown reaction Table 2-6.

Step	Temperature (°C)	Time	
Denaturation	95	5 minutes	
Denaturation	95	30 seconds	
Annealing	60	30 seconds	3 Cycles
Elongation	72	30 seconds	
Denaturation	95	30 seconds	
Annealing	58	30 seconds	3 Cycles
Elongation	72	30 seconds	
Denaturation	95	30 seconds	
Annealing	56	30 seconds	34Cycles
Elongation	72	30 seconds	
Final elongation	72	10 minutes	
Hold	10	∞	

Table 2-6: Single Colony PCR conditions

Following PCR amplification 5μl of SC-PCR product was run on an agarose gel, to confirm the inserted DNA was the correct size and that the colonies selected were of sufficient quality.

2.3.4.6 *Exosap DNA clean up*

Originally, the selected SC-PCR products were cleaned prior to sequencing by an exosap reaction, a process that uses Exonuclease I (*ExoI*) to digest residual single stranded DNA primers into single dNTPs, and alkaline phosphatase, which dephosphorylates individual dNTPs rendering them inactive.

The exosap reaction involved incubating 10µl SC-PCR product, with 1U FastAP alkaline phosphatase (Fermentas; Cat #: EF0651), 0.01U *ExoI* (New England Biolabs; Cat #: M0293S) and 1x FastAP buffer (from 10x stock containing supplied with *Exo1* and consisted of: 100mM Tris-HCL (pH8), 50mM magnesium chloride, 1M potassium chloride, 0.02% Triton X-100 and 1 mg/ml bovine serum albumin (BSA)) at 37 ˚C for 30 minutes, with an additional 20 minutes at 80 ˚C to inactivate the exosap enzymes. Following exosap inactivation, 5µl of the exosaped SC-PCR product was sequenced and precipitated via the methods described in section 2.5.2.

This exosap step was subsequently replaced by the more cost effective and efficient microclean DNA clean up step (section: 2.5.1).

2.4 Genomic sequencing

Genomic DNA was utilised for the study of candidate genes and identification of mutations within rare renal cancers (see chapter 5).

2.4.1 Whole exome sequencing (WES)

A total of between 3 and 5ug of stock genomic DNA at a concentration of 100ng/µl extracted from oncocytoma tumours was sent for Whole Exome Sequencing (WES) to Dr Michael Simpson (National Institute for Health Research, Guys and St Thomas' Foundation Trust, King's College London). In

summary, exons were captured and enriched via SureSelect All Exon v5 Target Enrichment System (Agilent; according to manufactories instructions), before massively parallel sequencing on an Illumina HiSeq2000 sequencer generating paired-end reads of approximately 100 bp.

Sequence reads obtained were computationally mapped via Novoalign (Novocraft Technologies) to the reference human genome (GRCh37/hg19). Analysis of coverage and depth of the sequencing was calculated via a combination of custom scripts and BedTools package (Code.google.com). Coverage of the gene was considered sufficient if more than 80% of the exon was sequenced at a read depth of greater than 20. Further analysis to identify and validate variations such as single nucleotide substitutions, small insertions or deletions between the sequencing and reference genome, was undertaken via in-house and SAMtools software package, and annotated via Annovar documentation tool. The frequency of variants, if they had been previously observed was annotated by comparison to the dbSNP132 and 1000 Genomes SNP calls databases (June 2011) and the in-house database of variants previously identified by exome sequencing of other samples by the same method.

The exome data was supplied in '.txt' format detailing all variants identified and additional information such as variant type, location, gene, and SNPdb reference for each sample after alignment to Human Genome (hg19/GRCh37). Raw data were supplied in '.BAM' and '.BAM.BAI' format and accessible via the Integrative genomics viewer (IGV (Robinson et al., 2011)). Files detailing the coding variants identified for each sample were also supplied. These files omitted all low quality

variants with a phred score of less than 20, corresponding to a 1% error rate of being falsely called. A phred score of 20 is widely considered the minimum threshold for sequencing quality (Guo et al., 2014; Koboldt et al., 2012). All variants located in the non-coding or intronic regions, were excluded, as these regions were not integrated to the same breadth, depth or quality as the exons. Bioinformatic analysis of the exome data to identify somatic oncocytoma variants was thus conducted on the remaining data.

2.4.2 Whole genome amplification (WGA)

To preserve DNA stocks, whole genome amplification (WGA) was conducted on stock genomic DNA prior to targeted whole gene screening via Sanger sequencing. Any variants identified from the WGA DNA were subsequently confirmed by retesting the stock genomic DNA. WGA was performed using REPLI-g Mini kit (Qiagen; Cat #: 150025) following manufacturers instructions. In summary, 50ng genomic DNA (5µl of 10ng/µl) was mixed with 5µl Buffer D1 (denaturation buffer) via vortexing then briefly centrifuged and incubated for 3 minutes at room temperature. The reaction was neutralised by the addition of 10µl Buffer N1 (neutralisation buffer), vortexed then briefly centrifuged. 30µl of WGA master mix (29µl REPLI-g Mini reaction Buffer and 1µl REPLI-g Mini DNA polymerase) was then added to the neutralised DNA sample resulting in a total volume of 50ul. The reaction was then incubated at 30℃ for 16 hours, before inactivation at 60℃ for 3 minutes. 1µl WGA was run on an agarose gel to confirm successful reaction, before being diluted 1:20 in nuclease-free water for use in Sanger Sequence gene screen. WGA DNA was stored at -20℃ until required.

2.4.3 Standard genomic PCR

To confirm the variants identified via WES and to screen candidate genes in additional samples, a targeted approach was undertaken. Two different PCR master mixes (outlined below) were utilised depending on the optimal mix identified during primer optimisation.

1: Roche master mix PCR reaction, used to amplify 20ng genomic DNA (stock or WGA), using the same Roche master mix described in the primary CoBRA reaction protocol (section:2.3.2.1)

2: Biomix Red PCR reaction, used to amplify 20ng genomic DNA (stock or WGA) The PCR master mix consisted of 12.5μl 2x Biomix Red (Bioline; Cat #: BIO-25005), 0.8μM of each primer and 5.5μl nuclease free water.

Reactions were either run on a standard PCR (std PCR (Table 2-7)) or using a touchdown PCR (Table 2-4) protocol depending on the primer optimisation. PCR products were run on agarose gels to confirm successful amplification. All primers and the optimal PCR conditions used to amplify and screen genomic DNA for mutations are listed in Table 7-2 and Table 7-3.

Step	Temperature ($^\circ$C)	Time	
Denaturation	95	5 minutes	
Denaturation	95	30 seconds	
Annealing	Ta	30 seconds	35 Cycles
Elongation	72	30 seconds	
Final elongation	72	5 minutes	
Hold	10	∞	

Table 2-7: Standard PCR conditions

2.5 <u>Sanger Sequencing</u>

Prior to running any sample on the Sanger sequencer the following preparation steps were undertaken for both genomic mutational study and methylation studies unless otherwise specified.

2.5.1 MicroCLEAN PCR clean up

To remove any residual primers and salts from the PCR product that could impede sequencing, PCR products were cleaned using the microCLEAN spin clean up protocol. Equal volumes of PCR product to microCLEAN (Microzone; Cat #: 2MCL-10) were added to a fresh plate (2.5μl of each was deemed to be the optimal for most PCR amplifications, although this volume was sometimes increased to 4μl if PCR product was weak when run on an agarose gel). Plates were centrifuged for 40 minutes at 2254 x g. Residual microCLEAN was then removed by further centrifugation of the plate upside down (without sealing film) on a tissue at less than 40 x g for 30 seconds. Samples were then ready for the addition of sequencing master mix (section: 2.5.2).

2.5.2 Sanger sequencing reaction

Sequencing reaction was conducted using Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems; Cat ♯: No: 4336917). A total volume of 10μl of sequencing reaction master mix was added to each well of microCLEANed PCR product. The sequencing reaction master mix contained 2μl 5x sequencing buffer, 0.5μl Big Dye, 0.5μM of either the forward or reverse PCR primer (one direction only per well) and 7μl nuclease free water.

For exosap cleaned PCR products, the nuclease-free water volumes were adjusted accordingly so the final volume per well, remained 10μl. Samples were then amplified on a PCR machine using the following sequencing reaction

protocol (Table 2-8). From this stage onwards it was important to keep samples in the dark to avoid degradation of the Big Dye.

Step	Temperature (°C)	Time	
Denaturation	95	5 minutes	
Denaturation	95	25seconds	
Annealing	Ta	25 seconds	29 Cycles
Elongation	72	4 minutes	
Hold	10	∞	

Table 2-8: Sequencing reaction conditions

2.5.3 Ethanol precipitation clean-up

Following the sequencing reaction, the samples were cleaned to purify the DNA and remove residual reagents. 2μl of 0.125 M sodium acetate and EDTA (Sigma-Aldrich; Cat ♯: E7889) was added to each sample, followed by 30μl of 100% ethanol (Sigma-Aldrich; Cat ♯:459844). Samples were pipetted up and down to thoroughly mix, and centrifuged for 20 minutes at 563 x g. Supernatant was removed from the pellet by a 30 second spin at 29 x g with the plate unsealed upside down upon a tissue. The pellet was re-suspended in 90μl of 70% ethanol by pipetting up and down before a further centrifugation step at 563 x g for a further 10 minutes, followed by removal of the supernatant by the same method as described above. Samples were then permitted to air dry for 5 minutes in the dark at room temperature to remove residual ethanol. Samples were re-suspended in 10μl HiDi Formamide (Applied Biosystems Cat ♯: 4311-329) and denatured for 5 minutes at 95°C before being sequenced via Sanger sequencing on 3730 DNA analyser (Applied Biosystems).

2.5.4 Loss of Heterozygosity

To establish if there was any LOH of *RASSF10* in sRCC samples (see chapter 4), a PCR reaction (see below) was conducted to analyse microsatellite markers flanking the *RASSF10* genomic location on genomic DNA from sRCC tumour and matched normal pairs. The microsatellites investigated had previously been identified by The West Midlands Regional NHS Genetics laboratory to possess high probability of heterozygosity (>0.7 probability (Table 2-9)).

Microsatellite marker ID	Genebank accession	Location	Distance to F10 (bp)		Prob. of het	Size range (bp)
			Start site	End Site		
D11S4116	Z52745	Chr 11: 12907174-12907502	124796	126479	0.78	205-224
D11S1334	Z24022	Chr 11: l2835634-12835775	196336	198019	0.81	134-150
D11S1794	Z51378	Chr 11: l3292332-13292701	-260362	-258679	0.84	249-273
D11S926>	Z17018	Chr 11: l3409034-13409183	-377064	-375381	0.75	135-145

	FORWARD	REVERSE	Ta F	Ta R
D11S4116	AACTGGTCCTTTAGACAGACAA	GAAAGCATCTCTATCTGCTGTG	59.4	61.2
D11S1334	TGCAGCATAGNCCTGT	AGCTTTATTGAAAGTCATTTTTG	47.9	49.5
D11S1794	GCTCCTAAAGGGTGGAGAC	TGCACTCAAAGCTGACAA	57.0	49.8
D11S926>	TGAGTGTGATGTATATGCTCATT	CAAGAGGTCAATTCTGGTGT	53.0	53.4

Table 2-9: LOH primer information (Chr 11 refers to chromosome 11, F the forward primer, R the reverse primer)

Each PCR amplification contained 20ng genomic DNA, 0.4µM microsatellite marker primers (Forward and Reverse) and 5µl 2x Biomix red made up to a total volume of 10µl per well with dH$_2$O. A standard PCR reaction protocol was used (Table 2-7) for a total of 28 cycles. Once complete, each PCR product was diluted and mixed with 120µl dH$_2$O. Separately, 5µl GeneScan 500 LIZ dye Standard (ThermoFisher Scientific; Cat #: 4322682) was mixed into 1 ml HiDi. In a fresh 96 well PCR plate 1µl diluted PCR product was added to 10µl HiDi/size standard mix, samples were denatured at 95 °C for 5 minutes then snap chilled on ice before running on Sanger Sequencer 3730 DNA analyser using Genescan run settings, and analysed using GeneScan Analysis software (Applied Bioscience). Comparison of tumour to matched normal permits the detection of changes in heterozygosity status for microsatellites. If microsatellites were identified to be heterozygous in normal and homozygous in tumour, it was deemed that LOH was present.

2.6 <u>Tissue culture</u>

All tissue culture (TC) was performed under sterile conditions in a microbiological safety cabinet class II (Thermo Electron Industries Holten LaminaAir). All media and reagents unless otherwise specified were warmed in a water bath to 37 °C. Prior to use, all surfaces and equipment was thoroughly cleaned with 70% ethanol to maintain sterile conditions. Any water used was double distilled (ddH$_2$O) and autoclaved prior to use. 1x phosphate buffered saline (PBS; OXOID; Cat #: BR0014G) was also autoclaved prior to use.

2.6.1 Cell culture maintenance

Cell lines were grown in 75 cm³ (T75) flasks (IBR stores; Cat #: IBR83.3911.002), in Dulbecco's Eagle Medium (DMEM; SIGMA; Cat #: D6429), enriched (complete DMEM) with 10% foetal bovine serum (FBS; Sigma; Cat #: F7524SIG), 1% L-Glutamine (L-Glut; Sigma; Cat #: G7513), 1% non-Essential amino acids (Sigma; Cat #: M7145), and 1% penicillin and streptomycin (100 units/ml penicillin, and 100μl/ml streptomycin (Pen/Strep); Gibco; Cat #: 15070-063) and incubated in a Sanyo CO_2 Incubator (model MCO-18AIC) at 37 °C with 5% CO_2.

2.6.2 Trypsination of cells

The following protocol was used to detach cells from the surface of the flask for counting, transfection or simple cell passaging. All media was aspirated from the flask prior to cells being washed gently in 10 ml PBS, which was subsequently aspirated from the flask and discarded. A volume of 0.25% Trypsin-EDTA (Gibco; Cat#: 25200-056) appropriate to the size of the flask or the plate (1ml for T75 flask, 20μl per well in 6 well plate), was gently applied to the growing side of the flask and gently swilled to ensure even coverage of the growing surface prior to incubation at 37 °C for 2-3 minutes or until cells could be seen to be detaching. Gentle agitation assisted detachment of the cells. Trypsin was quenched by the addition of complete DMEM to a ratio of 1:10 (i.e. 1 ml trypsin was quenched by the addition of 9ml media).

2.6.3 Counting cells

Trypsinised cells were counted using a haemocytometer prior to transfection. Suspended cells were placed in a 15ml falcon tube before 20μl of the cell suspension was added under the coverslip to each grid of a clean haemocytometer. Cells located within the grid were counted under a light

microscope; total cell count per ml was the average count across the two grids, multiplied by the haemocytometers conversion factor ($1\text{x}10^4$).

2.6.3.1 *Pelleting of the cells*

Following counting, the remaining cells in the falcon tube were centrifuged at 250 x g for 5 minutes, media was aspirated off. The cell pellets were then, either resuspended in complete media to obtain a desired concentration of cells per ml, or DNA/ RNA was extracted from the cell pellet (sections 2.6.5 and 2.6.6).

2.6.4 Treatment of cell lines with 5-aza-2'deoxycitidine

5-aza-2'deoxycitidne (5-Aza) treatment of the kidney cell lines (RCC11, UMRC3, A498, A704, KTCL140 and SKRC45) was conducted by Mr Dean Gentle prior to DNA or RNA extraction (sections 2.6.5 and 2.6.6). The DNMT inhibitor 5-aza-2'deoxycitidne was used to demethylate the genomic DNA of cell lines, permitting the analysis of gene expression for selected genes with and without methylation.

In summary, cell lines were grown to 30-50% confluency before starting treatment with 5-Aza. Once cell had reached the required confluency complete DMEM growth media was supplemented with 5μM 5-Aza (Sigma-Aldrich; Cat. No: A3656). The cell lines were then grown in this media for five days, with the supplemented growth media being replaced daily. For each cell line, a control non-treated cell line was grown simultaneously, in normal un-supplemented complete DMEM growth media.

After five days, cell lines were trypsinised and then pelleted for DNA or RNA extraction.

2.6.5 DNA extraction

DNA from cell lines was extraction by Mr Dean Gentle using the Qiagen DNeasy Blood and Tissue Kit (Qiagen; Cat #: 69506). Following the manufacturers' instructions for 'total DNA extraction from animal blood cells' (all reagents and spin columns were supplied within the kit unless otherwise mentioned); cell lines were pelleted (see 2.6.3.1) then re-suspended in 200μl PBS, 20μl proteinase K (Thermofisher scientific; Cat #: EO0491) and 200μl buffer AL before vortexing. The samples were then incubated at 56°C for 10 minutes, before the addition of 200μl 100% ethanol. Samples were subsequently transferred to a DNeasy Mini spin column and collection tube, and centrifuged at room temperature for 1 minute at 13,000 x g (centrifuge conditions remain the same for all subsequent centrifugation steps unless otherwise specified). The flow-through was discarded and 500μl buffer AW1 was added to the spin column. Samples were centrifuged again and flow-through was discarded. A fresh 2ml collection tube was then attached to the spin column, before the addition of 200μl buffer AW2 to the spin column. Samples were then centrifuged at room temperature for 3 minutes at 13,000 x g and flow-through was once again discarded. The collection tube was then replaced with a 1.5ml microcentrifuge tube and DNA was eluted from the membrane by the addition of 200μl of buffer AE, applied directly to the spin columns membrane. To increase the yield of DNA, samples were incubated at room temperature for 3 minutes, before a final centrifugation step at 13,000 x g for 1 minute. Extracted DNA was stored at -20°C until required.

2.6.6 RNA extraction

RNA was extracted from the 5-Aza DNA treated and control (untreated) kidney cell lines mentioned above (section 2.6.4) (Extraction was performed by D. Gentle).

In summary, cell pellets were homogenised by incubation for 5 minutes at room temperature in 1ml RNABee reagent (AMS Biotechnology; Cat #: CS-105B). Following this incubation 200μl chloroform (Sigma-Aldrich; Cat #: 288306) per 1ml RNA-Bee reagent was added and samples were shaken vigorously for 15-30 seconds before a further incubation on ice of 5 minutes. Samples were subsequently centrifuged at 12,000 x g for 5 minutes at 4˚C, this step separated the homogenate into 3 distinct phases; the lower phase consisting phenol-chloroform, an opaque white interphase and an upper colourless aqueous phase.

The RNA is located in the upper aqueous phase, which was carefully transferred into a fresh 1.5ml microcentrifuge tube. To precipitate the RNA, 500μl per1 ml RNA-Bee reagent was then added to the sample and mixed, before incubating for 5 – 10 minutes. Following incubation, samples were at 16,300 x g for 5 minutes at 4˚C, after which a visible pellet should be observed. Supernatant was then carefully removed and discarded taking care not to dislodge the pellet. The pellet was then subsequently washed by the addition of 1ml 75% ethanol, before vortexing and centrifuging at 7,500 x g for 5 minutes at 4˚C. supernatant was once again removed taking care not to dislodge the pellet formed. Samples were then permitted to air dry, before resuspension in 50μl RNase free water (Promega from IBR stores; Cat #: IBR-P1193PRO).

2.6.7 siRNA transfection

Small interfering RNA (siRNA) is commercially available, and an invaluable tool for transient knockdown of gene expression *in-vitro*. For siRNA knockdown of *RASSF10* (see chapter 4), the cell lines SKRC45 and KCTL140 were trypsinised, counted and seeded at $6x10^4$ cells per well in 6 well plates and replaced in the incubator for 24 hours to settle and re-attach. Following incubation, the media was removed and cells were washed with 3ml PBS, the PBS was aspirated before the addition of 2.29 ml fresh growing media. A siRNA master mix was created containing; 6.25 nM of siRNA either targeted for RASSF10: Hs_LOC644943_6 FlexiTube siRNA predesigned siRNA directed against human RASSF10 (NM_001080521, XM_928030; Qiagen; Cat#: SI04229617) or siRNA with no target: AllStars Negative Control siRNA (Qiagen: Cat# 1027280), mixed into 200µl serum-free OPTI-mem media (Gibco: Cat#: 31985-047). 10µl Interferin reagent (Polyplus; Cat#: 409-10) was then added to the siRNA/serum-free media solution and vortexed thoroughly before centrifuging down. The siRNA master mix was then added drop-wise to each well of the 6 well plate containing cells and incubated at 37 °C for 72 hours ready for protein isolation or soft agar assay. Final siRNA concentration was 2.5 nM per well. Each transfection was performed in triplicate.

2.6.8 Soft agar assay

Soft agar assays were employed to assess the ability of a cell line to undergo anchorage independent growth, and is deemed a physiologically representative study for cancer growth (Borowicz et al., 2014). The assay was conducted in 6 well plates, with 2 wells per transfection (transfections were repeated in triplicate (see section 2.6.4). The main set up, involved suspending cells in a low

density 0.35% agar, which is sandwiched between two layers of denser 0.7% agar.

This was achieved by the following protocol: All reagents and utensils were pre-warmed and stored in 42 °C water bath to prevent solidification until required. Firstly, 1.4% (w/v) Noble agar (BD biosciences; Cat #: 214220) was dissolved and superheated to sterilise in 100 ml sterile ddH$_2$O before cooling to 42°C. DMEM media was made up to the following concentrations required for the two agars by the dilution of 10x DMEM (Sigma-Aldrich; Cat#: D2429).

1x DMEM consisted of 10% 10x DMEM, 10% FBS, 5% of 7.5% Sodium bicarbonate (Sigma-Aldrich: Cat #: S5761), 1% Pen/Strep, 1% L-Glut and was made to a volume of 50 ml with ddH$_2$O.

2x DMEM contained 20% 10 x DMEM, 20% FBS, 10% of 7.5% Sodium bicarbonate, 2% Pen/Strep, 2% L-Glut and was made to a volume of 100 ml with ddH$_2$O. Mixing differing ratios of DMEM 1x and 2x with the nobel agar created the agar solutions (Table 2-10).

Reagents	0.7% Agar		0.35% Agar	
	Volumes (ml)	Final%	Volumes (ml)	Final%
1X DMEM	-	-	20	50
2X DMEM	75	50	10	25
1.4% Nobel Agar	75	50	10	25
	Total 150 ml	Ratio 1:1	Total 40 ml	Ratio 1:1:2

Table 2-10: Components and ratios for agars required for soft agar assay

To a 6 well plate, 2ml of the 0.75% agar (base layer) was added to each well and allowed to set at room temperature. Meanwhile transfected cells were trypsinised, counted and re-suspended in growth media to a concentration of 5×10^5 ml^{-1}, 100µl of cells (total 50,000 cells/well) was gently mixed with 1 ml 0.35% agar (slightly cooled) and added to each well on top of the set base agar. This was allowed to set before a final 2 ml 0.75% agar (top layer) was added and left to set. Plates were incubated at 37 °C in a humidified incubator for 28 days, and prevented from drying out by the addition of up to 100µl growth media when required. Colonies were counted and imaged under a light microscope when ≥30 cells (≥100µm) could be observed.

2.7 **Expression and Protein analysis**

2.7.1 **DNase treatment of RNA**

DNase treatment was performed to ensure all residual DNA was removed from an RNA sample before cDNA synthesis (this was particularly important for analysis for *RASSF10* a single exon gene, where due to the lack of exon:exon boundaries genomic DNA is not distinguishable from the RNA or cDNA).10µg RNA was gently mixed with 2U *DNase1* and 0.1 x volume 10X DNase I Buffer (both from Ambion DNAs free Kit; Thermo Fisher Scientific; Cat ♯: AM1906) and incubated for 20 minutes at 37 °C. The reaction was inactivated by the addition of 0.1 x volume DNase Inactivation Reagent (Ambion DNAs free Kit) and incubated for 2 minutes at room temperature with occasional mixing. Samples were centrifuged for 5 minutes at 10,000 x g before the supernatant containing purified RNA was removed to a fresh tube and stored at -80 °C until required.

2.7.2 cDNA synthesis

Complementary DNA (cDNA) was synthesised using Superscript III reverse transcriptase kit (Life Technologies; Cat ♯: 18080-093). In summary, 1µl 20 ug random primers (Promega; Cat ♯: C1181), 1µl 10 mM dNTP and 11µl RNase free water was added to 1µg purified RNA, and gently mixed, before incubation at 65°C for 5 minutes. Samples were then placed on ice for 1 minute, before the addition of 6µl a second master mix, which consisted 4µl 5x First-Strand Buffer, 1µl 0.1 M DTT, and 1U Superscript III Reverse Transcriptase (200 units/µl) bringing the total sample volume to 20µl. Samples were incubated on a PCR machine using the following reverse transcription PCR program (RT-PCR): 25°C for 5 minutes, 50°C for 60 minutes, before a further incubation step at 70°C for 15 minutes. cDNA was stored at -20 °C until required.

2.7.3 Protein isolation

To extract protein from cells, growth media was aspirated from cells and cells were gently washed with room temperature PBS that was then discarded. For a 6 well plate, 100µl RIPA lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 0.1%(v/v) sodium dodecyl sulphate (SDS) (Sigma-Aldrich; Cat ♯: 71727), 1mM EDTA; 0.5% (v/v) Na-deoxycholate, 1% (v/v) Igepal and '*cOmplete'* protease inhibitor cocktail (Roche; Cat #: 04693116001) which contained 15µg/ml Pancrease-extract, 1.5µg/ml Pronase, 0.8µg/ml Thermolysin, 1.5µg/ml Chymotrypsin, 0.2µg/ml Trypsin and 1 mg/ml Papain) was added to each well and incubated at room temperature for 5 minutes to allow cells to lyse. Cells were scraped with a cell scraper (IBR stores; Cat #: IBR-BC323) and collected into a 1.5 ml tube, before being placed on ice for a further 20 minutes. Samples were then

centrifuged at 16,300 x g at 4°C for 30 minutes, before carefully transferring the supernatant to a fresh 1.5 ml tube and stored at -80 °C until required.

2.7.4 Protein quantification

Protein concentration was measured using DC protein assay (Bio-Rad; Cat ♯:500-0113) following the manufactures instruction.

Firstly, a serial dilution in RIPA buffer of a protein standard was prepared from stock of 2mg/ml BSA standard (Pierce; Cat ♯: 23209), to achieve the final concentrations of 0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2 and 1.4µg/µl. Next mixture C was created by the addition of 20µl Reagent S per every ml required of reagent A. Then to a flat-bottomed 96 well plate, 5µl of sample protein or size standard was added (each sample and standard was run in duplicate), followed by 25µl mixture C finally 200µl Buffer B was added to each well. Samples were mixed gently to avoid generation of bubbles, and incubated for 20 minutes at room temperature to allow the development of the colour. The plate was then run on Wallac Victor3 fluorometer (Perkin Elmer), which measured the optical density at 690 nm wavelength of each sample. A standard curve was generated from the absorption recorded for the protein standards (Pearson's correlation r >0.9). The concentration of samples was then calculated using the regression equation derived from the standard curve.

2.7.5 Western blots

2.7.5.1 *SDS-PAGE*

Proteins were separated according to their size, via Sodium dodecyl sulphate polyacrylamide gel (SDS-PAGE) electrophoresis. The SDS detergent is vital for

denaturing the protein and applying a negative charge to the peptide, thus allowing progression through the gel when subjected to an electrical current.

Acrylamide gels were prepared with using 12% resolution gel (12% (v/v) Acrylamide Protogel (30% stock National Diagnostics; Cat ♯: ED-201), 375 mM TRIS pH 8, 0.1% (v/v) SDS and polymerized with 10% (w/v) Ammonium persulfate (APS) and 5μl Tetramethylethylenediamine (TEMED; Sigma-Aldrich; Cat ♯: T9281). The gel was poured into glass gel plates, locked in a gel stand, up to 4/5 full. Water was gently added to the final fifth to ensure all bubbles on top of the resolution gel were removed. Once set, water was tipped off and the gel dried. A stacking gel (3% (v/v) Acrylamide Protogel, 50 mM TRIS pH 6.8, 0.1% (v/v) SDS with 10% (w/v) APS and 5μl (v/v) TEMED) was poured in the final fifth of the gel plates on top of the resolution gel and a loading comb mould was inserted.

20μg of sample protein was mixed with 5x loading buffer (0.313 M Tris HCL pH 6.8, 10% (v/v) SDS, 0.05% (w/v) Bromophenol blue, 50% (v/v) glycerol). Samples were denatured at 95 °C for 5 minutes, then snap chilled on ice prior to loading into the stacking gel. Gels were inserted into a gel tank, and submerged in 1 x SDS running buffer (25mM Tris, 192mM glycine, 0.1% (v/v) SDS, pH 8.3; Geneflow; Cat ♯:B9-0032)). Samples were run next to Pageruler Plus Pre-stained Protein Ladder (ThermoFisher/ Fermentas; Cat ♯: SM1811FER) at 100 V until loading dye had reached the resolution gel, then at 120 V for 60 minutes or until loading dye reached the bottom of the resolution gel.

2.7.5.2 *Wet Transfer*

To transfer protein from the acrylamide gel to a polyvinylidene fluoride membrane (PVDF; GE Healthcare; Cat ♯: 10600021), a sponge and 3 layers of 3mm thick Whatman chromatography paper sheets (IBR stores; Cat ♯: IBR-WHATMAN3) were saturated with transfer buffer (25 mM Tris, 192 mM Glycine and 20% MeOH; Geneflow; Cat ♯:B9-0056), before being placed in an open transfer cassette. Bubbles were removed by rolling over the top Whatman paper with a strippette. The PVDF membrane was cut to the size of the gel and activated by submersion in 100% methanol before being placed on top of the soaked Whatman paper and sponge with air bubbles carefully removed. Gel plates were then gently opened, and the stacking gel was discarded. The remaining resolution gel was placed on top of the membrane, and a further 3 sheets of saturated Whatman paper and another saturated sponge was placed on the top of the gel, bubbles were gently removed again and the cassette was closed and locked, creating a sandwich of sponge, 3 Whatman papers, membrane, Gel, 3 Whatman papers and sponge. Cassettes were place in the transfer insert and replaced into a gel tank, filled with transfer buffer and an ice block. It was vital to ensure that the cassette was loaded with the gel closest to the cathode and the membrane further away. An electrical field was applied at 100 V for 90 minutes.

Once the transfer was complete, the membrane was removed and blocked by submersion in blocking solution (5% (w/v) milk powder (Marvel form IBR stores Cat #: IBR-MARVEL) in PBS with 0.1% (v/v) Tween (Sigma-Aldrich; Cat #: P2287) overnight at 4 °C.

2.7.5.3 *Immunodetection*

To visualise/ identify proteins of interest; blocked membranes were first probed with an antibody specific for the protein of interest . Primary antibodies were diluted to their optimal concentration in 5 ml blocking solution and incubated on a roller at room temperature for the optimised time. Rabbit-α-RASSF10 (previously made in-house as reported by V K Hill et al., 2011) was diluted 1:1000 and incubated for 4 hours, while mouse-α-Tubulin (house-keeping gene to ensure equal loading; Sigma; Cat #: T9026) was diluted 1:20,000 and incubated for 1 hour.

Following incubation membranes were placed in a tray on a shaker and thoroughly washed three times in 10 ml PBS with 0.1% Tween, leaving the PBS/Tween solution on the membranes for 5 minutes between each wash to remove residual unbound antibody.

Due to the primary antibodies being unconjugated, a secondary antibody step was required. Membranes were therefore reincubated with 5 ml blocking solution and 1:20,000 dilution of appropriate secondary antibody (α-Rabbit IgG-HRP (Sigma; Cat ♯: RABHRP2) or α-mouse IgG-HRP (Sigma Cat ♯: RABHRP1) on a roller for 1 hour at room temperature. Membranes were then washed thoroughly 3 times for 10 minute in 10 ml PBS with 0.1% tween on a shaker. In order to visualise bound antibodies, membranes were developed using Pierce ECL Plus enhanced chemiluminescence (ECL) solution (ThermoFisher; Cat ♯: 32132). ECL solutions were mixed 40:1, solution A:B (approximately 1ml total required per membrane). Membranes were then placed on saran wrap (IBR stores; Cat ♯: IBR-SARAN), covered with the complete ECL mixture and left to incubate at room

temperature for 5 minutes. Post incubation, membranes were removed from the ECL solution, wrapped in fresh saran wrap and placed in a developing cassette. Medical X-Ray film (Kodak: Cat ♯: 525 3349) was then exposed to the membrane for as long as was required to detect a signal and film was developed using a SRX-101A medical film processor (Konica Minolta). A positive signal was identified as a black band on the X-ray film and protein size was established by comparison to the protein size standard ladder.

2.7.5.4 *Quantification of Western blots*

In order to quantify the signal strength of the western blot, and ascertain protein abundance, the developed western blot film was imaged and bands were digitally quantified using the Gene Tool software (as used to image agarose gels). Band intensity was computationally calculated as the intensity of the band minus the intensity of the background film.

2.7.5.5 *Stripping membrane*

In order to re-blot membranes to detect housekeeping genes, membranes were stripped of all antibodies by boiling for 5 minutes. Membranes then required re-blocking with blocking solution overnight before incubation with antibodies as before (see section 2.7.5.3).

2.7.6 Mass spectrometry

A preliminary mass spectrometry (MS) study of a stable HEK293 cell line overexpressing RASSF10-Flag under different cell cycle synchronisation conditions was conducted by Dr. David Matallanas and team (Systems Biology Ireland, University College Dublin). The cell cycle synchronisation conditions investigated were as follows (Rosner et al., 2013):

- *Control*: un-transfected HEK293 included to rule out contaminates

- *Growth*: HEK293-RASSF10-Flag under normal growing conditions

- *Noco 1x*: HEK293-RASSF10-Flag incubated with 100 ng/ml Nocodazole for 21 hours, arresting cell cycle at G2/M phase

- *Thy-1x*: HEK293-RASSF10-Flag incubated with 2 mM Thymidine for 21 hours, blocking cell cycle at G1/S

- *Starved*: HEK293-RASSF10-Flag incubated in serum free media for 16 hours, cells arrested in G0/G1 phase.

Each condition was repeated in triplicate and run twice on the mass spectrometer resulting in 6 datasets for each condition.

In brief, cells were grown under the appropriate conditions before being lysed as described in section 2.7.3. Lysates were then immunoprecipitated (IP) with Protein G beads and antibodies specific to Flag. IPs were digested into peptides and centrifuged. The supernatant was eluted, alkylated to break di-sulfite bonds cleaned and enriched to remove all salts via spin column/tip before loading and running on LTQ OrbiTrap Mass Spectrometer (Thermo Scientific). Output files were analysed using MaxQuant (Version 1.3) and label free quantification intensity (LFQ) was calculated. Raw spectrometry data was exported to Microsoft Excel 2011, where the average output for each cell cycle synchronisation condition, the ratio between conditions, and the significance between conditions were calculated.

2.8 <u>Bioinformatic Analysis</u>

2.8.1 Human Genome Assembly

Human genome assembly build Hg19/GRCh37 was used as the reference genome for all comparisons, primer designs and UCSC and Ensemble referencing throughout this project.

2.8.2 The Cancer Genome Atlas (TCGA) and cBioPortal

The TCGA (Tcga-data.nci.nih.gov, 2015) is a multi-network collaboration funded by National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI), with the aim of interrogating the genetic constitutes of multiple cancers via multiple high throughput platforms and making the data publically available. To date, data from multiple platforms including clinical information, DNA methylation, somatic mutations, copy number variant, mRNA sequencing and gene expression is publically available for 34 cancer types in more than 11,000 samples. DNA methylation data (Level 3 from HumanMethylation450) was downloaded for analysis in chapter 3 and chapter 4 (Downloaded: 02/10/2012). For the study in chapter 3, all 81 pRCC samples (Samples ID listed in Table 7-4) available at the time of download and 100 ccRCC (Samples ID listed in Table 7-5), (randomly selected via random number generator) were downloaded. For chapter 4, methylation data of 160 ccRCC tumours and matched normal was downloaded from the TCGA (Samples ID listed in Table 7-6) (Downloaded: 01/03/2013). Summary of the clinical data for the TCGA ccRCC and pRCC samples downloaded is displayed in Table 2-11 In chapter 5 variants identified by exome sequencing were compared to TCGA data using cBioPortal (cbioportal.org, 2015; Gao et al., 2013) . cBioPortal provides a user friendly, accessible interpretation of all mutations and copy number

alterations compiled from TCGA data and other publications, for each gene across different cancers.

Histology	Gender	Average Age (Years)	UICC Stage (2002)	TNM Grade			Average survival time (days)	Inclusion in project:		
				T	N	M		Chapter 3	Chapter 4	Chapter 5
ccRCC	F = 67 M = 132	62.2 (29 - 90)	I = 80 II = 18 III = 58 IV = 43	T1a = 55 T1b = 27 T2 = 25 T3a = 52 T3b = 29 T3c = 1 T4 = 8	N0 = 86 N1 = 5 n.d = 108	M0 = 158 M1 = 41	801 (18 - 2803) n.d = 117	100	160	-
pRCC	F = 25 M = 56	60.4 (31 - 83) n.d = 3	I = 45 II = 3 III = 22 IV = 8 n.d = 3	T1a = 37 T1b = 11 T2 = 3 T3a = 17 T3b = 7 T4 = 1	N0 = 19 N1 = 8 N2 = 4 n.d = 50	M0 = 47 M1 = 4 n.d = 30	804 (240 - 1556) n.d = 69	80	-	-

Table 2-11: Clinical data TCGA samples used in this study
Table summarises the clinical data for the TCGA samples used within this study. The table details tumour type: ccRCC, clear cell RCC and pRCC: papillary RCC patients age at diagnosis, gender of patient (**M** = male, **F** = female), the chapter in which the samples were used. Whether the samples were tumours (**T**) only, or if matched normal (**N**) were available. If applicable, the grades of the tumour are displayed according to the 2002 UICC staging classification (stage I-IV) and the TNM scoring (see section 1.4.1). Patient survival time in days is also documented. For all cases **n.d** = no data available.

2.8.3 Genomic structural, functional and variant information

2.8.3.1 *UCSC Genome browser*

University of California Santa Cruz (UCSC) genome browser (genome-euro.ucsc.edu) provides publically accessible information on genomic sequence, CpG island location , expression, regulatory features, ENCODE functional element locations, common SNPs and 450K humanbead methylation probe location. DNA sequences used for primer design were downloaded via UCSC DNA view tool.

2.8.3.2 *Ensembl*

The Ensembl genome browser (Ensembl.org) is a publically accessible database, detailing among other things genomic sequence, exon boundaries, transcripts, known genetic variations, SNPs and FASTA sequence for multiple species. Data are collected from multiple sources including but not limited to dbSNP, 1000 genomes project, Exome variant server, COSMIC and individual observations. Information regarding minor allele frequency (MAF) and variant prevalence across different populations was obtained from Ensembl and used to help identify potential variants of interest when analysing WES data (see chapter 5).

2.8.3.3 *Genecards*

Genecards (Genecards.org) is an integrative database providing an overview of individual gene functions, expression and proteomics compiled from over 100 sources. Genecards was used to obtain a foundation understanding of a genes function and expression levels in the kidney. This information was used to assist in the identification of potential gene of interest (GOI).

2.8.3.4 *Cosmic*

The Catalogue of Somatic Mutations In Cancer (COSMIC) (Cancer.sanger.ac.uk) provided detailed information for a multitude of cancers and genes, including the prevalence, distribution and somatic status of mutations identified in tumours. Additionally COSIMC also provides information regarding the methylation status and expression of genes within cancers although this data is mostly pooled from the TCGA datasets.

2.8.3.5 *1000 genome project*

The 1000 genomes (1000genomes.org) project was established to generate a catalogue of 'normal' human genetic variations from different non-diseased populations. The aim was to sequence greater than 2,500 genomes at a minimum of 4x coverage. To date a total of 2,577 samples have been sequenced and published from 5 distinct ancestry groups (East Asian= 523, South Asian= 494, African= 691, European= 514, and American=335) with a total of 26 subpopulations. In particular, for the study in chapter 5, 105 Japanese (Tokyo) samples have been screened (57 male and 48 female).

2.8.3.6 *Exome variant server (EVS)*

The NHLBI Grand Opportunity Exome Sequencing Project (ESP) Exome variant server (EVS) (Evs.gs.washington.edu) was established to generate a database of whole exome sequenced European American and African American populations, from samples with diverse phenotypes, with the aim of identifying genetic variants associated with heart, lung and blood disorders. To date a total of 6503 samples (2203 African-Americans and 4300 European-Americans) have been sequenced and variant frequency has been published for control and diseased samples, although individual phenotype data is not publically available.

2.8.3.7 *The Exome Aggregation Consortium (ExAC)*

ExAC is a database of exonic variants compiled from the whole exome sequences of over 60,700 unrelated individuals. All sequencing files were processed through the same pipeline to ensure consistency in the data. (Exac.broadinstitute.org, 2015)

2.8.3.8 *Japanese specific databases*

In Chapter 5, ethnic specific variant analysis was undertaken. Two databases were employed to assess the prevalence of specific variants associated with a Japanese population.

1. Japanese Single Nucleotide Polymorphisms (JSNP) (Snp.ims.u-tokyo.ac.jp; Hirakawa, 2002). A dataset detailing frequent variations in a Japanese cohort. To date 197,195 variations have been characterised, with allele frequencies obtained from a pool of 752 unrelated Japanese samples.

2. Integrated Japanese Project, (Nagasaki et al., 2015; Ijgvd.megabank.tohoku.ac.jp, 2015) which provided frequency data for over 4,300,000 single nucleotide variants (SNV) from 1070 Japanese individuals aged between 20 and 80 years.

2.8.4 Data analysis software

2.8.4.1 *CpG Viewer*

CpG viewer is an integrated tool used to visualise and analyse the methylation status of bisulfite modified DNA after single colony sequencing. Providing the user a clear output of the methylation state of each CpG di-nucleotide interrogated, by comparing single colony sequencing data to a reference sequence (Carr et al., 2007).

2.8.4.1 *Clustering analysis*

Clustering analysis conducted in chapter 3 was achieved via use of Opensource Cluster v3.0 (de Hoon et al., 2004). Supervised hierarchical clustering using Euclidean distance was performed using Cluster v3.0 on the 450K data. The output file was then visualised using Java tree view Version 1.1.6 (Saldanha, 2004) .

2.8.4.2 *Ingenuity Pathway Analysis (IPA)*

QIAGEN's Ingenuity Pathway Analysis (IPA)⍰ (ingenuity.com) software was employed in Chapter 3 to determine if any of the genes identified to be hyper or hypomethylated in specific RCC subtypes were associated with known networks, biological processes, functions, interactions, pathways or diseases.

2.8.4.3 *String db*

STRING database (Jensen et al., 2009; String-db.org, 2015) is a web based resource, which was used in Chapters 4 and 5, to visualise known and predicted protein-protein interactions (both physical and functional). Additionally, string db generates protein network diagrams, and highlights proteins known to be associated with disease and key pathways.

2.8.4.4 *Data for Annotation, Visualisation and Integrated Discovery (DAVID)*

DAVID Bioinformatics Resources 6.7 (Huang et al., 2009) was a tool that was used in all studies under taken in this thesis. DAVID uses multiple sources to group lists of genes according to pathways, known biological functions, or protein domains, which the genes have been associated with or encode. Permitting the identification of common features and functions from a large gene

list. Additionally, DAVID can provide information on diseases and which genes from the list have previously been associated.

2.8.4.5 *Integrative Genomics Viewer (IGV)*

The Integrative Genomics Viewer (IGV) (Robinson et al., 2011) is a high performance tool used for the visualisation and exploration of the large raw BAM and BAM.BAI files produced via the whole exome sequencing. IGV displayed all reads mapped to the exons exonic regions of the reference genome Hg19 (Figure 2 1 a), with quality and depth for each base displayed in a quality histogram. Deletions and single nucleotide variants were highlighted in the coverage histogram (either as a drop in histogram peak height (Deletion) or as two colours representing the two possible variants in the recorded proportions). Variants were also visible when analysing individual reads, with a deletion shown as gap in the read joined by a line, and a SNV identified as an the alternate nucleotide marked on the read (2 1 b-d). IGV was used in chapter 5 to assess the read depth, read quality and coverage of genes of interest and as a method to validate the somatic status of SNVs bioinformatically. By comparing the number of reads presenting with the variant to the number of wild type (wt) reads. Variants were considered somatic if present in greater than 15% of tumour reads and less than 5 %% of normal reads (threshold was set allowing forto rule out false positive reads).
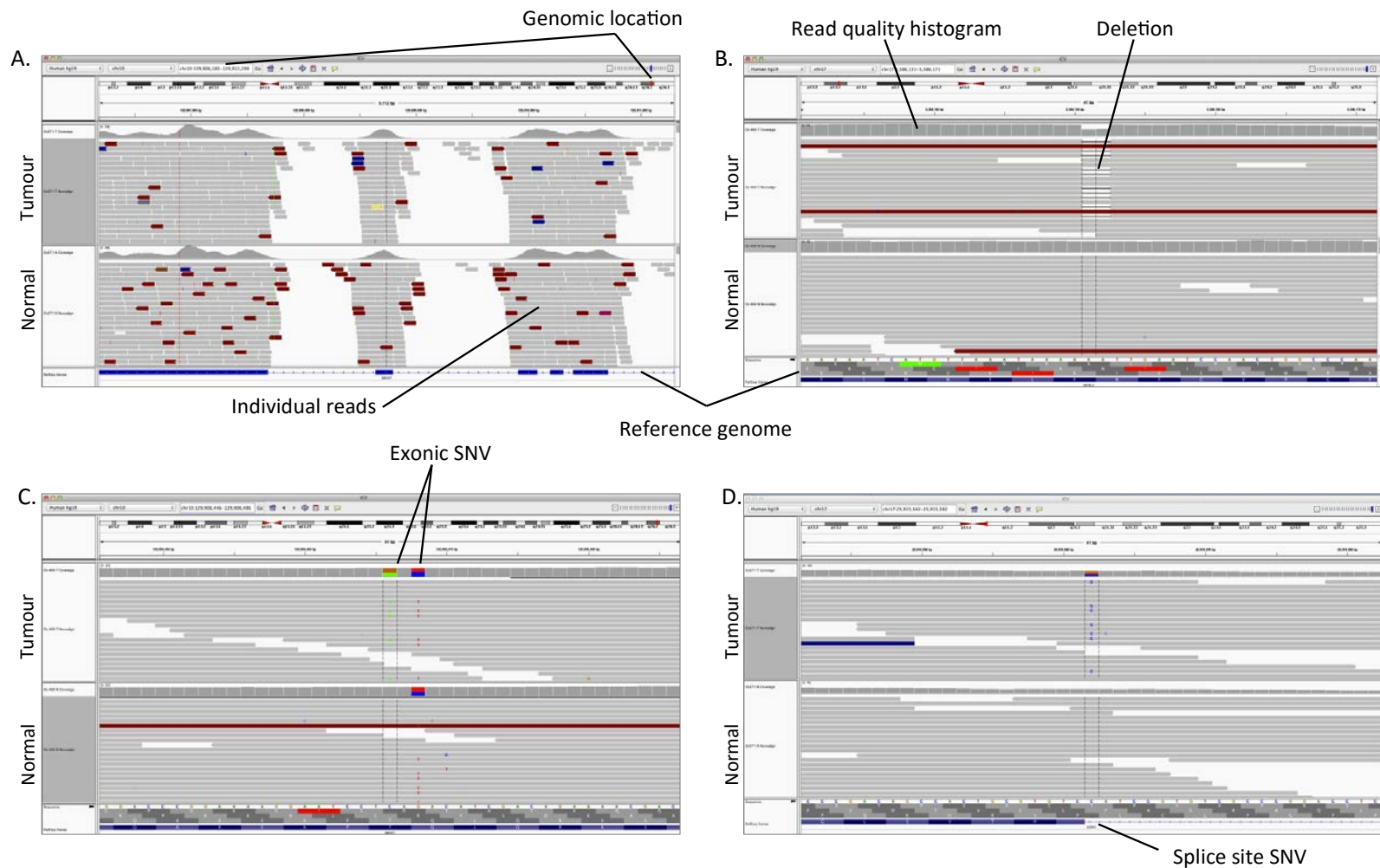
Figure 2-1: Example of IGV display output.
Screen shots of the IGV display. **A)** Zoomed out view displaying the coverage and read depth of WES mapped to exons (Blue squares in reference genome). **B)** Example of a somatic deletion of 'AA' dinucleotide (represented by grey line and drop in histogram quality) in *DERL2* (variant: T=42.2% and N=0% of reads). **C)** Exonic SNVs are displayed in individual reads and quality histogram (colours denote the proportion of each nucleotide). Two variants are shown for *MIK67*, a somatic SNV (variant: T=41.4% and N=0% of reads) and a germline SNP (rs4750936 MAF:0.23(T)). **D)** Somatic SNV in the exonic splice site region (variant: T=36.2% and N=0% of reads), identified in *KSR1* in an intronic location with close proximity to the exon boundary.

2.8.4.6 *Mutation surveyor*

Softgenetics Mutation Surveyor DNA variant analysis software (Softgenetics.com, 2015) was used in Chapter 5 to analyse Sanger sequencing data produced when screening genes for recurrent mutations. Multiple sequencing files can be uploaded and analysed, with mutation surveyor producing a report on the sequencing quality and highlighting any regions where the sequence deviates from reference for manual inspection. Regions of lower sequencing quality were also investigated manually to confirm if a true variant could be detected or whether only background noise or sequencing artefact was observed.

2.8.4.7 *Viewing electropherograms*

Sanger sequencing electropherograms were viewed either by 4Peaks (Nucleobytes.com, 2015) for Mac, or Chromas (softpedia, 2015) for Windows.

2.8.4.8 *Prediction of the functional effects of SNV's in Chapter 5*

Online bioinformatics tools were utilised to aid the identification of potentially pathogenic SNV and splice site variants. The tools described below were employed in chapter 5, however it should be noted that the following tools provide a prediction of pathogenicity only, therefore tools were used in conjunction with each other to support a prediction.

2.8.4.8.1 *Polyphen-2*

Polymorphism Phenotyping v2 (PolyPhen-2)(Adzhubei et al., 2010) is a web based tool designed to assess the probability of a missense variant being deleterious. An iterative algorithm that compares the mutant amino acid change to the wild type amino acid

sequence and assesses a total of 11 criteria (eight sequence based and three structural) to predict if an amino acid substitution is detrimental. Scores are given between 0, representing a predicted benign substitution, and 1, predicted likely to be damaging.

2.8.4.8.2 *Sorting Tolerant From Intolerant (SIFT)*

SIFT (Kumar et al., 2009) was the second prediction tool used to analyse missense variants. Similar to Polyphen-2, SIFT provides predictions as to whether an amino-acid substitution would have a deleterious impact. However, SIFT differs from Polyphen in the approach undertaken to predict the effect of a variant. By assessing the conservation of an amino acid throughout evolution, SIFT predicts whether a variant is predicted damaging (with a score of 0) or tolerated (with a score of 1).

2.8.4.8.3 *Splice site variant prediction*

To assess whether variants located within the splice site region were likely to result in loss of the splice site, the *in-silico* prediction tool Berkeley Drosophila Genome Project: Splice Site Prediction by Neural Network (Fruitfly.org, 2015) was used. By using the NNSPLICE 0.9 version (1997) Fruitfly.org was able to predict, with a 5% false positive rate, if a splice site was maintained when a SNV was present.

2.8.5 Statistical analysis

The data and statistical analyses were conducted using the following software Microsoft Excel (2011), IBM SPSS v21.0 (2012) and Graphpad Prism v5.0a (2007). Statistical tests included student t-test to test for significance between two groups with similar variance, Fisher's exact test (small cohorts) and Chi-squared test (larger cohorts). These were

performed using Graphpad Prism and were used to compute significance from a contingency table of categorical data. Significance was taken as P<0.05.

When analysing multiple groups of unequal sample size one-way ANOVA with Game-Howell post hoc test was used to test for significance. This was performed using SPSS, with significance once again set as P<0.05.

Statistical tests such as the Student t-test possess a 5% false detection rate, therefore to rule out genes (chapter 3) or proteins (chapter 4) falsely deemed significant, a false detection rate (FDR) correction was applied if more than 20 t-tests were conducted. The pre-calculated t-test P-values were ranked from smallest to largest (*Rank*), and the FDR was calculated according to Equation 5:

$$FDR = \left(\frac{P\ value\ x\ Total\ tests}{Rank} \right)$$

Equation 5: FDR formula
Statistical investigation into patient survival was performed in Graphpad Prism using Kaplan-Meier survival curves and statistical comparisons assessed via Mantel-Cox test. A P<0.05 was deemed significant.

Chapter Three: **DNA methylation profiles of oncocytoma and chromophobe RCC**

3.1 <u>Introduction</u>

3.1.1 Techniques to analyse DNA methylation

The study of DNA methylation can be achieved by several different techniques from analysing the DNA methylation of individual genes to the DNA methylation of the whole genome. Some of the most commonly used techniques are described below.

3.1.1.1 *Bisulfite modification*

As DNA methylation is not maintained during PCR amplification most investigation techniques used to analyse DNA methylation requires the DNA to be permanently modified to permit the identification of methylated CpG dinucleotides after PCR amplification.

To overcome this limitation, DNA can be bisulfite modified. A method first described in the 1970's but not generally used until the early 1990's, bisulfite modification has become a foundation technique in analysing DNA methylation (Frommer et al., 1992; Hayatsu, 2008). Bisulfite modification involves chemically converting unmethylated cytosine residues into uracil, which are subsequently replaced with thymidine during PCR amplification. This occurs through the process of sulfonation (the addition of a bisulphite group) followed by a deamination step to remove of an amine group. Finally sulphur trioxide is removed by desulfonation, i.e. the addition of hydroxide anion (Figure 3-1a).
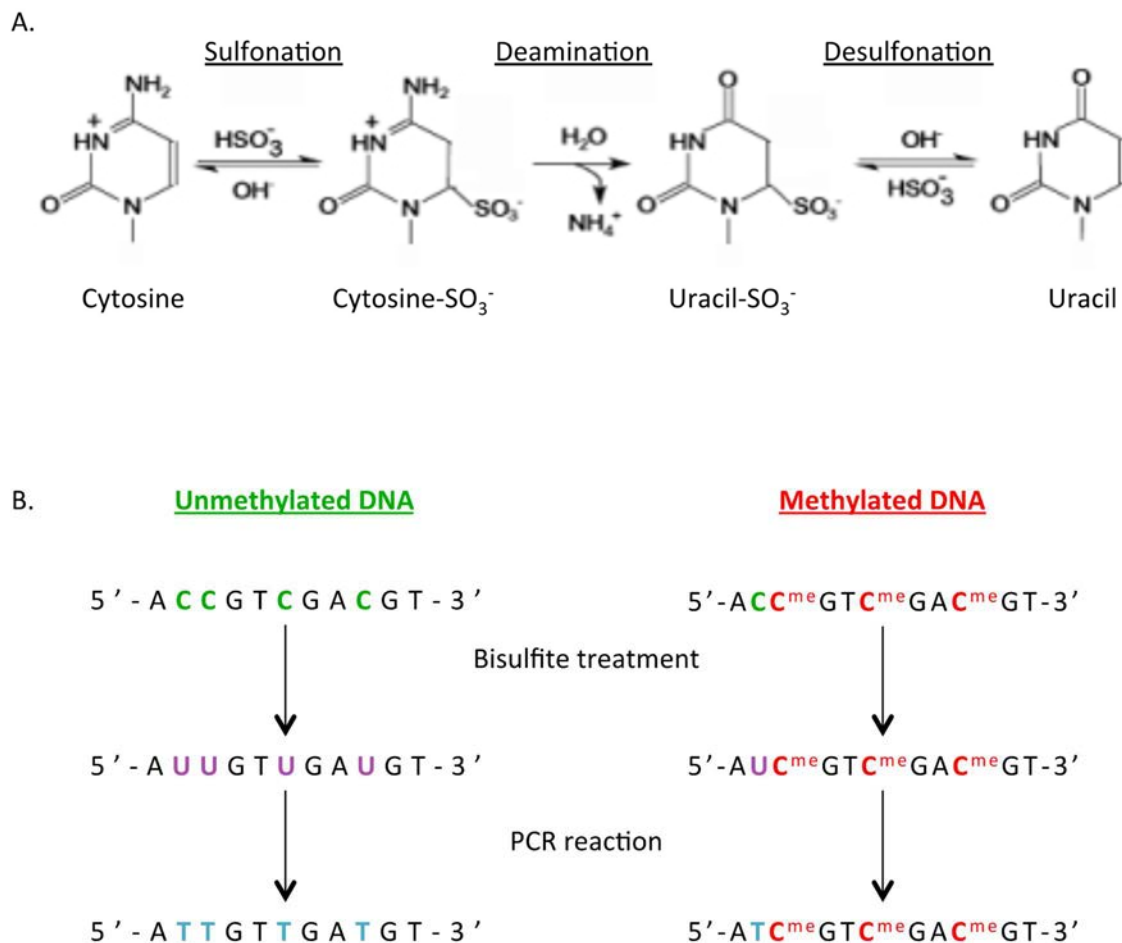
Figure 3-1: Bisulfite modification of DNA
A) The chemical reaction that takes place during bisulfite modification, converting unmethylated cytosine to uracil. (Image from Activemotif.com, 2015). **B)** Schematic of a DNA sequence and steps undertaken to bisulfite modify DNA. Image shows pre- and post-bisulfite modification and following the first PCR reaction if the cytosine residues are unmethylated (Green) or methylated (red). Blue 'T' nucleotides are converted unmethylated cytosine's (adapted from Abcam.com, 2015)

Following this conversion of cytosine to uracil, PCR based experiments can be used to ascertain DNA methylation status, as all unmethylated cytosine residue will have been converted to thymidine while methylated 5-methylcytosine residues remain intact (Figure 3-1b)(Activemotif.com, 2015; Hayatsu, 2008). An alternative technique to investigate methylation include Methylated DNA immunoprecipitation (MeDIP), where antibodies specific to methylated DNA are used to pull down and immunoprecipitate DNA which is methylated prior to real time PCR, sequencing and micro array analysis (Brena and Plass, 2009;Activemotif.com, 2015; Sims et al., 2014).

3.1.1.2 *Methylation Specific PCR (MSP)*

MSP is a sensitive PCR method to determine the methylation status of candidate genes of interest. By using two sets of primers, one specific for the methylated DNA sequence (MSP) and a second set that specifically amplifies the unmethylated DNA sequence (USP unmethylated specific PCR), each sample is investigated twice to obtain the complete picture of the methylation status. Successful amplification of a PCR product is indicative of the region of interest being methylated (when using MSP primers) or unmethylated (when using USP primer) (Herman et al., 1996; Heyn and Esteller, 2012).

As this technique directly interrogates the methylation status of the DNA to which the primers bind, it can offer a more sensitive approach to detecting methylation than CoBRA (see 3.1.1.3). The specificity this technique provides makes it a very useful tool when investigating regions of known high methylation density such as the CpG island of *RASSF1A* (Herman et al., 1996; Heyn and Esteller, 2012).

3.1.1.3 *Combined bisulfite restriction analysis (COBRA)*

CoBRA is another PCR based methylation detection technique, first reported by Xiong and Laird, 1997 that is now frequently used. CoBRA is an effective, rapid and simple PCR based method to analyse the DNA methylation status of regions of interest. Compared to MSP, CoBRA is a more quantitative test, providing an in insight into the overall level of methylation at a region of interest, a feature that is partially useful for investigating novel regions where the DNA methylation state is uncharacterised (Heyn and Esteller, 2012; Xiong and Laird, 1997)Bisulfite DNA is PCR amplified using primers that are designed to omit bias for methylated or unmethylated DNA, by containing a 50:50 T/C split for each possible methylated cytosine residue. Following amplification, DNA is then digested using enzymes such as *BstUI*, which recognises and cleaves 5'-CG|CG-3', or *Taqα1* which specifically cleaves 5'-TC|GA-3'. If DNA is unmethylated these sites are lost during the bisulfite modification becoming 5'-TGTG-3' or 5'-TTGA-3' respectively. Digested PCR products are run on a gel next to undigested PCR product. A positive result for methylation results in multiple bands being visible in the digested column.

3.1.1.4 *Quantitative methylation techniques*

The above techniques identify if DNA methylation is present, however, the accuracy at which they are able to quantify the level of DNA methylation, is limited. The development of quantitative methylation techniques such as Methylight (discussed in section 4.1.1), quantitative MSP (which uses a fluorescent probe to quantify DNA amplification in real time, Brena and Plass, 2009; Trinh et al., 2001), DNA methylation arrays (discussed below 3.1.2) and bisulfite modified DNA sequencing, have since

addressed this limitation. Methylight and quantitative MSP are both PCR based methods that can provided quantitative information into the overall level of DNA methylation for a region of interest. But do not provided information as to the status of DNA methylation for each individual CpG locus within that region, unlike DNA methylation arrays or bisulfite modified DNA sequencing (Brena and Plass, 2009; Trinh et al., 2001).

DNA sequencing of bisulfite modified DNA can be conducted by candidate gene Sanger sequencing, pyrosequencing or via sequencing the whole genome on a next generation sequencing platform, a technique known as whole genome bisulfite sequencing (WGBS) (Li and Tollefsbol, 2011). All of these techniques provided information about the methylation status of each individual CpG locus within the genome or region of interest investigated (Brena and Plass, 2009; Li and Tollefsbol, 2011) . Pyrosequencing is a sequence by synthesis technique conducted on bisulfite modified DNA, which records and measures the fluorescence emitted as of specific biotinylated nucleotides are incorporated into the newly synthesised DNA strand. As these biotinylated nucleotides are incorporated into the DNA, fluorescence is generated by a luciferase based enzyme cascade, and quantified. By releasing the biotinylated nucleotides into the sequencing reaction in a predetermined order, researchers are able to identify the sequence of the bisulfite modified DNA and therefore determine the proportion of C or T dinucleotides at a given CpG locus, thus providing a quantitative assessment of the methylation status of individual CpG locus (Brena and Plass, 2009; Tost and Gut, 2007). Within this study however DNA methylation was assessed using the latest quantitative DNA methylation Array (section 3.1.2)

3.1.2 DNA Methylation Arrays

DNA methylation arrays provide another method to quantitatively analyse the DNA methylation status of individual CpG dinucleotides, in a high-throughput and cost-effective manner. DNA methylation arrays were originally adapted from SNP arrays, to identify DNA methylation in bisulfite modified DNA (section 3.1.2.1). Since their development, they have become an increasingly popular tool for assessing global methylation (Goldengate BeadChip array 2006, Bibikova et al., 2006)).

The Illumina Goldengate assay was the first array platform released to investigate DNA methylation. It contained 1,536 CpG probes (Figure 3-2) that mapped to 807 cancer associated genes. A total of 71.4% (577) of the cancer associated genes investigated by the Illumina Goldengate assay had their DNA methylation status determined 2 or more probes on the array (Bibikova et al., 2006; Illumina, 2006). The Goldengate array has revolutionised the study, and our understanding of DNA methylation. Permitting the identification of distinct patterns of hypermethylation when comparing between different tumour types, including the identification of possible CIMP+ phenotypes (McRonald et al., 2009; Wolff et al., 2010), and identifying associations between the prevalence of DNA hypermethylation and the invasiveness of tumours. As was reported by Wolff et al when comparing the DNA methylation profiles of different types of urothelial cancers and non-cancerous bladder tissue (Wolff et al., 2010).
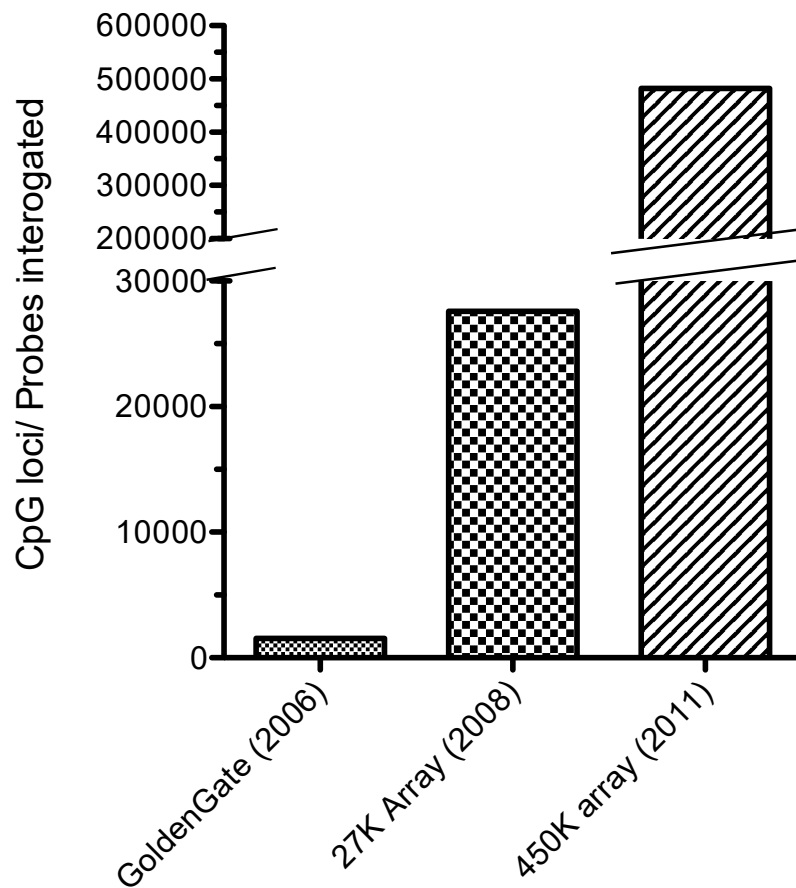
Figure 3-2: Illumina DNA methylation array comparison
The total number of CpG loci interrogated by each Illumina DNA methylation array and the year the technology was commercially available. GoldenGate refers to the GoldenGate BeadChip array 27K array = Infinium HumanMethylation27 BeadChip, and the 450K array = Infinium HumanMethylation450 BeadChip.

Additionally, the Goldengate array permitted the identification of differentially methylated genes in different tumour types. For example a study in to pRCC and ccRCC methylation identified *HOXA11* and *HOXC6* to be preferentially methylated in pRCC when compared to ccRCC (McRonald et al., 2009). Another study coupling the Goldengate array with expression data, aimed to identify novel methylated genes that may contribute to tumourigenesis (Ricketts et al., 2013).

The next stage in methylation array evolution was the release of the Infinium HumanMethylation 27 BeadChip (27K array), which interrogates the methylation status of 27,578 CpG loci (Figure 3-2) across 14,495 genes (Weisenberger et al., 2008). This array provided researchers with a greater breadth of information as to the DNA methylation status of the epigenome with studies identifying frequent methylation of novel genes not previously associated with cancer, such as *OVOL1* (methylated in 40% ccRCC samples) and *SST* (methylated in 31% ccRCC samples) in ccRCC (Ricketts et al., 2012). Other studies using this array have identified differentially methylated genes associated with clinical features, prognosis and survival. For example a study by Hill et al into DNA methylation of breast cancer identified 6 notable genes (*RECK, SFRP2, UAP1L1, ACADL, ITR,* and *UGT3A1*) to significantly reduce patient survival time when methylated (Hill et al., 2011). With further studies having identified novel frequently methylated genes that may be used as biomarkers to distinguish between oestrogen receptor (ER) positive (21 methylated genes) and ER negative (6 methylated genes) breast cancer subtypes (Fackler et al., 2011).

3.1.2.1 *Infinium HumanMethylation450 BeadChip*

The latest DNA methylation array, the Infinium HumanMethylation450 BeadChip (450K array), provides even greater genome coverage, interrogating the methylation status of 482,421 CpG loci (Figure 3-2). This means that the methylation status of >99% of Refseq genes and 96% of CpG islands (CGIs) is investigated by one array (Bibikova et al., 2011; Illumina, 2012). Of the probes incorporated in the array >99.3% are targeted for CpG dinucleotides (Sandoval et al., 2011).

The principle behind the 450K array is similar to the Goldengate array the 27K array. All three platforms identify C to T bisulfite conversion through quantitative genotyping by utilising an adapted SNP detection system. For the Goldengate and the 27K arrays this was achieved through the sole use of the Infinium I probe (Bibikova et al., 2006; Illumina, 2006; Weisenberger et al., 2008). For the 450K array however, two probe technologies are used; Infinium I and Infinium II. Infinium I probes account for 51.7% of the probes which map to CpG islands and only 15.9% of the probes which map to regions not associated to a CGI (>2000bp from CGI) (Dedeurwaerder et al., 2011).

The two probes differ in the method by which methylation is detected (Figure 3-3); Infinium I probes (Figure 3-3a) require two beads for each CpG locus integrated. One bead specific for the methylated CpG locus and the second specific for the unmethylated CpG locus. If the sample sequence is complimentary to the probe's methylation status for a CpG locus ('CA' unmethylated probe sequence, 'CG' methylated probe sequence), single base extension can occur and the incorporation the fluorescently labelled nucleotide is detected at the next nucleotide past the CpG locus.

A.
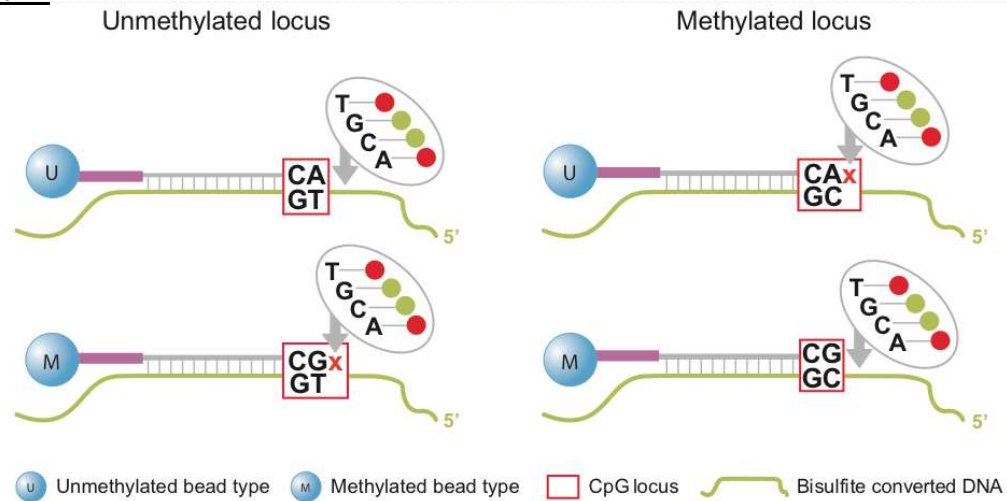
B.
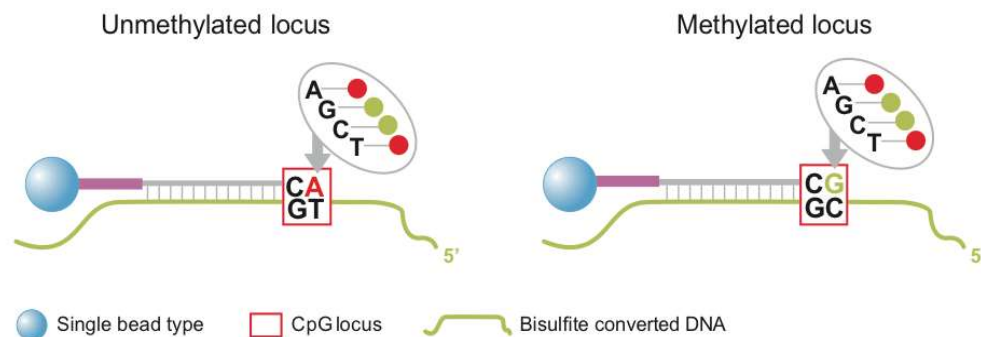
Infinium II

Figure 3-3: Infinium probe design used in the Illumina DNA methylation arrays
A) Schematic of the Infinium I probe used in the GoldenGate BeadChip, Infinium HumanMethylation27 BeadChip, and Infinium HumanMethylation450 BeadChip. Two separate probes are used to measure methylated (M) and unmethylated (U) CpG loci through recording the florescence to the neighbouring nucleotide as it is incorporated. If sequence DNA does not match the reference probe elongation is terminated and no florescence is recorded. **B)** Schematic of the Infinium II probe used in the Infinium HumanMethylation450 BeadChip. One bead containing probes for both methylated and unmethylated CpG locus is used. Methylation is determined by identifying the specific colour emitted as the fluorescent-labelled nucleotide is incorporated at the CpG loci. (Image: Illumina, 2012).

At sequences not complimentary to the probes methylation status for the interrogated the CpG locus, no extension is observed and thus no fluorescent detected allowing identification of the methylation state of that particular locus (Bibikova et al., 2011; Illumina, 2012).

The Infinium II probe technology, (Figure 3-3b) uses only 1 bead containing probes specific for either the methylated sequence or the unmethylated sequence and interrogates the actual 'C' base in question of the CpG locus rather than the base just after. To differentiate between the methylated and unmethylated sequences the Infinium II technique relies on the detection and differentiation of the colour of the fluorescence emitted when the fluorescent nucleotides are incorporated (red=A or T and green=C or G nucleotide) (Bibikova et al., 2011; Illumina, 2012). The quantity of methylation at a CpG locus is displayed as a β-value. β-values are calculated by comparing the fluorescence recorded for methylation specific probes to that of unmethylated probes (section 2.3.3).

The incorporation of two different techniques within one array, has not been without controversy, with studies showing variation in the range of β-values recorded, with Infinium II being less sensitive at detecting extreme methylation values (β-values close to 0 and 1) (Dedeurwaerder et al., 2011). Thus it is now widely accepted that normalisation of the two probe types should be conducted prior to analysis.

The distribution of the 450K arrays CpG probes throughout the genome has been characterised and defined according to their location in relation to Refseq genes and CpG islands (CGIs). Firstly, in relation to the CG rich CGIs (Figure 3-4a),
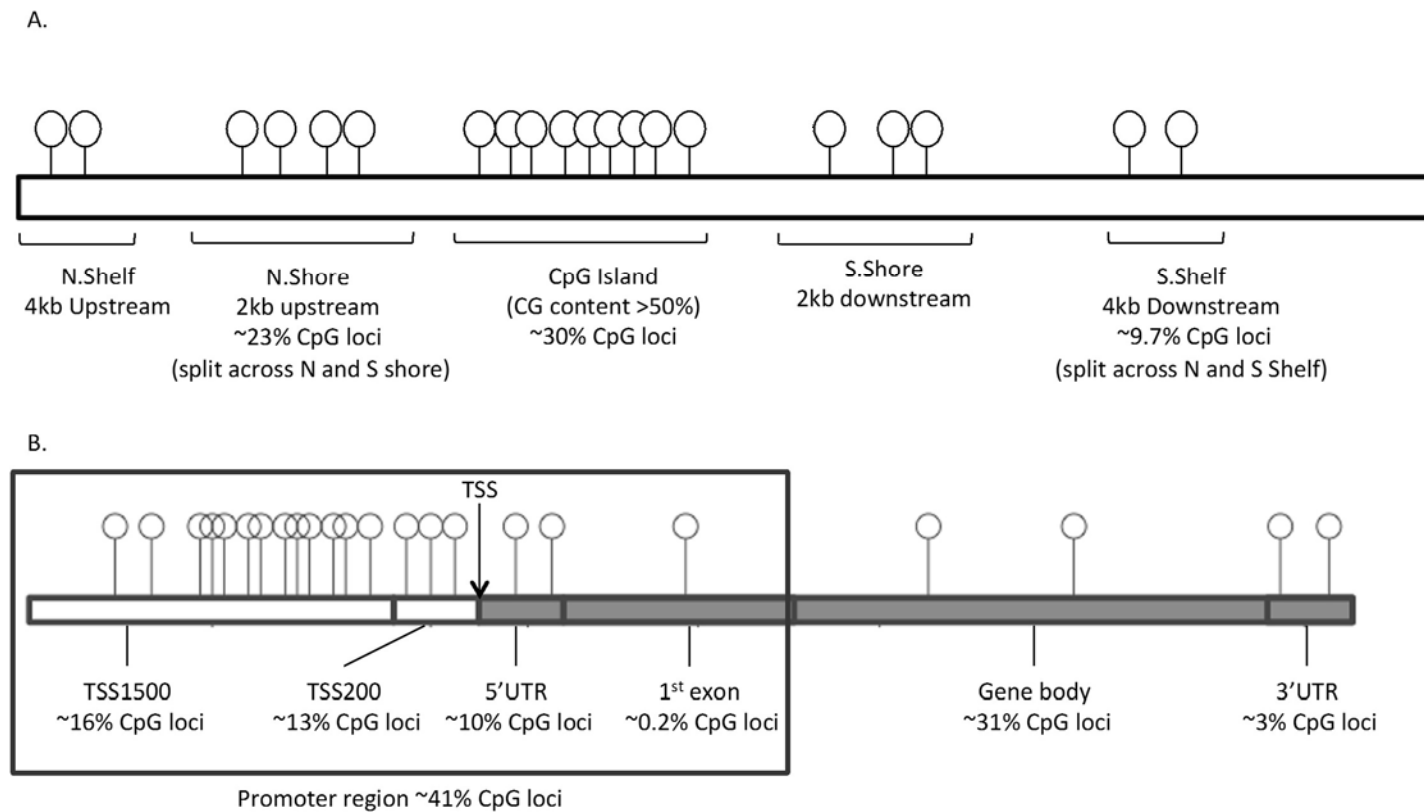
Figure 3-4: Distribution of the Infinium HumanMethylation450 BeadChip CpG loci

Schematic diagram showing the prevalence and distribution of the CpG loci investigated by the Infinium HumanMethylation450 BeadChip array. CpG loci (white circles) are shown in relation to; **A)** CpG islands and **B)** Refseq genes. Abbreviations N: North S: South TSS: transcriptions start site, UTR: untranslated region. Image adapted from Bibikova et al., 2011

30.9% of 450K CpG probes map to the CpG islands (CG content >50%), 23% map to CpG shores (CG rich regions less dense than CGI, located less than 2kb up or down stream of the CGI), 9.7% map to CpG Shelves (less dense CG content than the shores and around 4kb up or downstream of the CGI) and 36.3% probes identified to reside in isolated CG dinucleotides referred to as 'open seas' (Bibikova et al., 2011; Sandoval et al., 2011). When mapped to Refseq genes, 74.4% of the total 450K CpG probes were associated with coding genes and classical mRNAs, 0.85% of total 450K probes were mapped to non-coding RNA (microRNA and long non coding RNA) and 24.6% of probes did not map to known transcripts, thus were defined as intergenic regions (Dedeurwaerder et al., 2011; Sandoval et al., 2011).

Additionally, probes were further characterised depending on where in a gene the CpG loci were located (Figure 3-4b); 41.3% of all CpG probes were identified to be located in the promoter region of genes (this was defined as all probes located within 1500bp of the Transcription start site (TSS) TSS1500, 200bp upstream of TSS (TSS200) 5' untranslated region (5' UTR) and the 1st exon), 3.16% of probes mapped to 3'UTR, 30.9% to the gene body (intronic regions and exons except 1st exon) and as previously mentioned, 24.6% mapped to intragenic regions (Dedeurwaerder et al., 2011; Sandoval et al., 2011).

The depth of information provided by the 450K array has led to a plethora of findings and studies. It is the predominant technology used by the TCGA consortium to analyse genomic methylation profiles of different cancers and has lead to the ability to compare across arrays, particularly with expression data. Promoting the identification of frequently hypermethylated genes which have been shown to inversely correlate with

gene expression, and elude to a particular methylation signatures which inhibits gene expression (Davis et al., 2014; TCGA, 2016, 2014, 2013, 2008). The 450K array has since been used to identify patterns of methylation and methylated genes in a plethora of different cancers. Some examples of which include *HER*2+ breast cancer (Lindqvist et al., 2014) and cervical cancer (Lando et al., 2015). In addition to studies attempting to differentiate between tumours subtypes such as recurrent and non-recurrent chordoma tumours (Alholle et al., 2015), or different breast cancer subtypes (Stefansson et al., 2015) based on the DNA methylation patterns. Other studies have associated DNA methylation with pharmaceutical response, for example a study by Trabelsi et al, identified that methylation of *MGMT* in glioblastoma improves drug efficiency in chemotherapy, by making the tumour more sensitive to the drug (Trabelsi et al., 2015). Not all studies using the 450K array focus on cancer, with studies using the array to characterise variation within a population and to assess environmental effects on DNA methylation profiles (Argos, 2015; Busche et al., 2015; Heyn et al., 2013).

3.1.3 Study's Aim

As described in section 1.4.2.1.3 onwards, the malignant chRCC and the benign renal oncocytoma account for 3-5% and 5% of renal neoplasms respectively (Lopez-Beltran et al., 2010). However, these two subtypes are both histologically and visually (when imaged by medical imaging techniques) similar making diagnosis of the subtypes challenging even for experts (Baert and Sartor, 2006; Millet et al., 2011; Stec et al., 2009; Vera-Badillo et al., 2012). Although cytogenetic and immunohistological differences have been reported in chRCC and oncocytomas the use of these features to aid diagnosis

lacks reliability and specificity (Ng et al., 2014; Yusenko, 2010), leaving histological interpretation of renal biopsies and medical imaging the current approved method of diagnosis (Ng et al., 2014).

Despite the similar histologies and appearances, these two RCC subtypes have different patterns of progression (with chRCC being malignant and oncocytoma being benign) and therefore have different courses of treatment. chRCC is predominantly treated via surgery (Vera-Badillo et al., 2012), while the course of treatment for oncocytoma is usually one of surveillance with surgical intervention as a lest resort if the tumour become symptomatic (Neuzillet et al., 2005). As a result, correct diagnosis and identification of the RCC subtype is vital for efficient and effective treatment. Further investigations into potential distinguishing features could provide the key to reliably differentiating these two subtypes. One such feature could be DNA methylation; having been previously been proposed as a potential biomarker to identify cancer type, prognosis and drug sensitivity in several other cancers (section 1.3.4), (Exner et al., 2015; Heyn and Esteller, 2012; Trabelsi et al., 2015).

Though studies have previously been conducted to assess and compare the methylation profiles of different RCC subtypes including chRCC and oncocytoma, these have predominantly focused on identifying and comparing the methylation status of candidate cancer associated genes, previously reported to be methylated in RCC or other cancers, such as *VHL, RASSF1A, FHIT* and *MGMT* (Costa et al., 2007; Dulaimi et al., 2004; Maher, 2013; Morris et al., 2010). However, of the genes investigated no differential methylation was reported for chRCC and oncocytoma, this observation was attributed to

chRCC and oncocytoma possessing similar histology's and deriving from the same cell type (Costa et al., 2007; Dulaimi et al., 2004).

Other studies focusing on the analysis and characterisation of genome wide methylation profiles of RCC have primarily focused upon ccRCC (Lasseigne et al., 2014; Liu et al., 2014; McRonald et al., 2009; Morris et al., 2011; Ricketts et al., 2014, 2013, 2012). Although the analysis had yet to be published by the onset of our study, the TCGA consortium had also begun screening the methylation profiles of chRCC (TCGA, 2013) and pRCC (TCGA, 2016) with the aim of profiling the genetic, methylation and expression profiles of the two RCC subtypes. At the time of formulating the aims for this study (September 2011) only one report had been published investigating the global DNA methylation profile of chRCC and renal oncocytoma. Here it was observed that although the DNA methylation profile was mostly similar between the two histologies, small regions of variation were identified. However, this study was conducted using BACS clones and the results where not related to genes which my be deregulated (Arai et al., 2011). The limited studies conducted into the methylation profiles of chRCC and renal oncocytoma mean that the DNA methylation profile of the two histologies is poorly understood. We therefore propose (see objective 1 (section 1.7)):

- To investigate and characterise the global DNA methylation patterns of chRCC and renal oncocytoma using the high throughput Illumina Infinium HumanMethylation450 BeadChip array.
- Identify differential patterns of methylation and methylated genes that may provide markers by which the chromophobe and oncocytomas histologies may be distinguished.

129

- Compare our findings with the methylation data available from the TCGA in order to identify methylated genes or patterns unique to chRCC or oncocytoma.

3.2 <u>Results</u>

3.2.1 Global methylation analysis of oncocytoma and chromophobe RCC

The following work was published in Epigenetics, March 2013 (Slater et al., 2013).

3.2.1.1 *Infinium HumanMethylation 450 BeadChip*

A total of 4 non-diseased normal kidney samples, 21 oncocytoma (1 sample run in duplicate) and 20 chRCC (Table 2-1) were outsourced for bisulfite modification and analysis on the Infinium HumanMethylation 450 BeadChip (as described in section 2.3.3).

3.2.1.2 *Validation of the array*

At the time of this study, the 450K array was a relatively novel assay with previous genome wide methylation studies being conducted on the 27K array (Arai et al., 2012; Ricketts et al., 2013). Initial validation of the array focused on confirming the reproducibility and reliability of the data. To investigate this, one oncocytoma was sent in duplicate to be run on the 450K array. The correlation between the two replicates was determined, by comparing the β value recorded for all CpG loci between the two replicates and statistically tested. Pearson's correlation coefficient identified a strong linear correlation (R=0.997), confirming the reproducibility of the array (Figure 3-5).
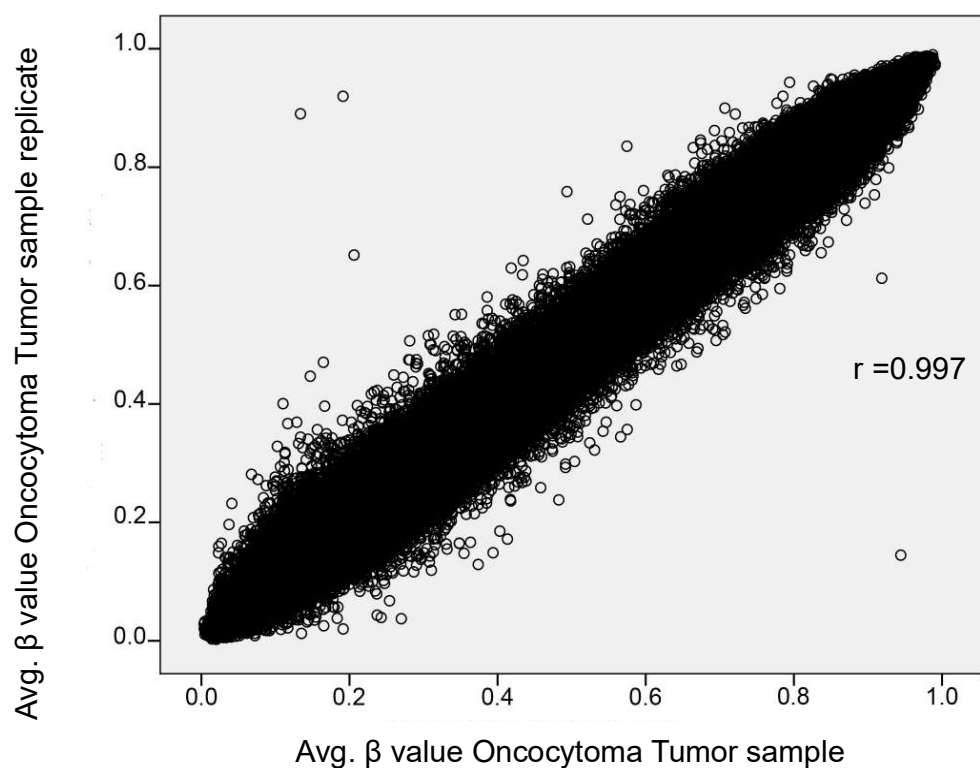
Figure 3-5: Validation of the reproducibility of the 450K array
To validate that the 450k is a reliable and reproducible array, one oncocytoma was run on the array in duplicate. β-values for all CpG loci were plotted for each replicate and the correlation was calculated by Pearson's correlation coefficient giving the values r=0.997. This is indicative of high level of correlation between the two run thus confirming the arrays reproducibility.

To confirm the quality of the data and to ascertain how the β value actually corresponds to the level of DNA methylation present, single colony sequencing of bisulfite modified DNA (section 2.3.4) for three candidate genes (*BNC1*, *FOXL1* and *OVOL1*), previously identified to be frequently methylated in ccRCC (Ricketts et al., 2012) was conducted. For each CpG locus within the sequenced region, that was interrogated by the 450K array, the methylation index (MI) (the percentage of methylated CpGs identified by single colony sequencing) was calculated and compared to the β value obtained form the 450K array (Figure 3-6 to Figure 3-8). Of the loci investigated, we identified that when the β value was >0.5, the methylation index for the same CpG dinucleotide was >50% indicating the presence of methylation, while low β values of <0.25 typically corresponded to a MI for the same dinucleotide of <0.25. For this study, a β value of >0.5 was considered indicative of the presence of methylation at a given locus. This threshold is consistent with the 450K validation reports (Dedeurwaerder et al., 2011; Illumina, 2012; Weisenberger et al., 2008).

BNC1: NM_001717

450K Probes:
cg02713760| cg07016258| cg06523224
cg17124224| cg26654798| cg26429925

CpG island:188

Oc:5

β value: 0.46/ 0.49/ 0.82/ 0.78/ 0.65/ 0.29

MI: 50%/ 50%/ 100%/ 100%/ 80%/ 20%

Oc:8

β value: 0.23/ 0.58/ 0.42/ 0.29/ 0.52/ 0.57

MI: 10%/ 80%/ 80%/ 40%/ 100%/ 90%

Ch:8

β value: 0.75/ 0.87/ 0.9/ 0.82/ 0.86/ 0.77

MI: 100%/ 100%/ 100% 100%/ 90%/ 100%

Ch:9

β value: 0.64/ 0.5/ 0.82/ 0.62/ 0.54/ 0.4

MI: 80%/ 100%/ 80%/ 60%/ 50%/ 50%

Oc:3

β value: 0.47/ 0.17/ 0.16/ 0.07/ 0.15/ 0.12

MI: 55.5%/ 22.2%/ 0%/ 11.1%/ 11.1%/ 0%

Oc:18

β value: 0.24/ 0.25/ 0.24/ 0.08/ 0.23/ 0.12

MI: 14.3%/ 0%/ 42.8%/ 14.3%/ 14.3%/ 0%

Ch:13

β value: 0.26/ 0.49/ 0.29/ 0.05/ 0.39/ 0.17

MI: 12.5%/ 62.5%/ 37.5%/ 0%/ 25%/ 0%

Ch:15

β value: 0.23/ 0.2/ 0.05/ 0.03/ 0.1/ 0.27

MI: 14.3%/ 14.3%/ 0%/ 14.3%/ 14.3%/ 14.3%

Oncocytoma   **Chromophobe**

Figure 3-6: Validation of the 450K array: *BNC1*

Validation of β values by bisulfite single colony sequencing of *BNC1* (previously reported as frequently methylated in ccRCC (Ricketts et al., 2012)). Encompassing CpG loci cg02713760 (south shore), cg07016258, cg06523224, cg17124224, cg26654798 and cg26429925, located in the TSS1500. Five of the CpG loci investigated resided within the *BNC1* CpG Island. Eight samples (4 chRCC and 4 oncocytoma) with a range of β values for the aforementioned CpG loci were selected. Hypermethylated loci (β value > 0.5) presented with a high methylation index (MI > 50%), while low β value loci correlated with a low MI.

Diagram shows schematic for part of the gene (blue) and CpG island (green) with region interrogated by bisulfite sequencing marked by the dotted line. CpG loci are represented by the red marks on gene schematic and red arrows on the bisulfite clone sequencing readout. Bisulfite sequencing rows represent individual alleles, with each circle indicating location and methylation state of a CpG locus (black circle: methylated; white circle: unmethylated). Ch denotes chromophobe samples, Oc oncocytoma samples and MI the methylation index.

FOXL1: NM_005250

CpG island:154

450K Probes:
cg00624737 | cg07748217

Oc:5
β value: 0.55/ 0.76
MI: 44.4% 88.9%

Oc:18
β value: 0.53/ 0.82
MI: 40%/ 87.5%

Ch:8
β value: 0.88/ 0.87
MI: 87.5%/ 100%

Ch:14
β value: 0.56/ 0.83
MI: 57.1%/ 85.7%

Oc:13
β value: 0.19/ 0.25
MI: 11.1%/ 22.2%

Oc:3
β value: 0.32/ 0.28
MI: 16.6%/ 16.6%

Ch:13
β value: 0.2/ 0.19
MI: 0%/ 0%

Ch:17
β value: 0.19/ 0.24
MI: 0%/ 16%

Oncocytoma

Chromophobe
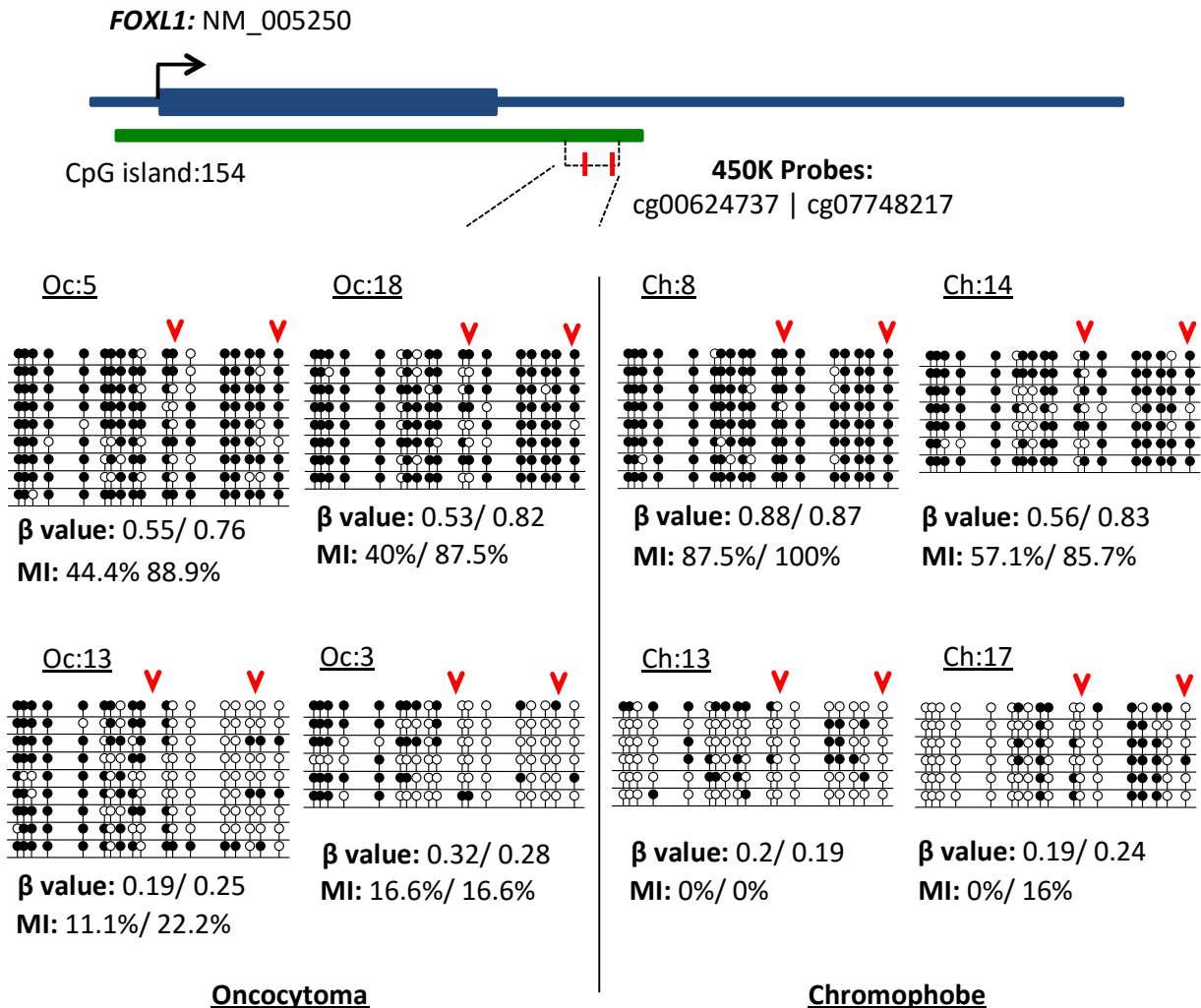
Figure 3-7: Validation of the 450K array: *FOXI1*

Validation of β values using bisulfite single colony sequencing of *FOXI1* (previously reported as frequently methylated in ccRCC (Ricketts et al., 2012)). Encompassing CpG loci cg00624737 and cg07748217 located within the 3'UTR and the CpG island. Eight samples (4chRCC and 4 oncocytoma) with a range of β values for the aforementioned CpG loci were selected. Hypermethylated loci (β value > 0.5) presented with a high methylation index (MI > 50%), while low β value loci correlated with a low MI.

Diagram shows schematic of the gene (blue) and CpG island (green) with region interrogated by bisulfite sequencing marked by the dotted line. CpG loci are represented by the red marks on gene schematic and red arrows on the bisulfite clone sequencing readout. Bisulfite sequencing rows represent individual alleles, with each circle indicating location and methylation state of a CpG locus (black circle: methylated; white circle: unmethylated). Ch denotes chromophobe samples, Oc oncocytoma samples and MI the methylation index

Figure 3-8: Validation of the 450K array: *OVOL1*

Validation of β values using bisulfite clone sequencing of *OVOL1* (previously reported as frequently methylated in ccRCC (Ricketts et al., 2012)). Encompassing CpG loci cg11569703 located within the gene body and the south shore. Eight samples (4chRCC and 4 oncocytoma) with a range of β values for the aforementioned CpG loci were selected.. Hypermethylated loci (β value > 0.5) presented with a high methylation index (MI > 50%), while low β value loci correlated with a low MI.

Diagram shows schematic of the gene (blue) and CpG island (green) with region interrogated by bisulfite sequencing marked by the dotted line. CpG loci are represented by the red marks on gene schematic and red arrows on the bisulfite clone sequencing readout. Bisulfite sequencing rows represent individual alleles, with each circle indicating location and methylation state of a CpG locus (black circle: methylated; white circle: unmethylated). Ch denotes chromophobe samples, Oc oncocytoma samples and MI the methylation index

3.2.1.3 *Identification of aberrant cancer specific methylation*

The 450K produces a large comprehensive dataset detailing the genome wide methylation profile for multiple samples. However, to identify cancer specific methylation, bioinformatic analysis was undertaken to remove technical and biological bias and generate a workable dataset. Raw data was supplied as .txt file detailing probe information, average β value and detection P value for each probe and sample. The pipeline we employed to filter the data is visualised in Figure 3-9. The first step was to assess the reliability of each probe, this was determined by analysis of the detection P-value. The detection P value was calculated by comparing the intensity recorded for each probe to an internal negative control. Multiple factors can affect the detection P value such as weak hybridisation of the DNA sample to the probe, low read intensity, and mutations within the DNA samples (Dedeurwaerder et al., 2011). Probes that did not comply with the reliability criteria, defined as a detection P value of <0.01 (Infinium, 2012; Sandoval et al., 2011) in all samples were omitted, 5607 probes were removed at this step.

The next stage was to omit gender bias, as epigenetic X- chromosome inactivation within females and single X chromosomes in males, can skew the methylation profile (Chen et al., 2013). This was achieved by the removal of all probes located on the X or Y chromosome, of which 11199 probes were found.
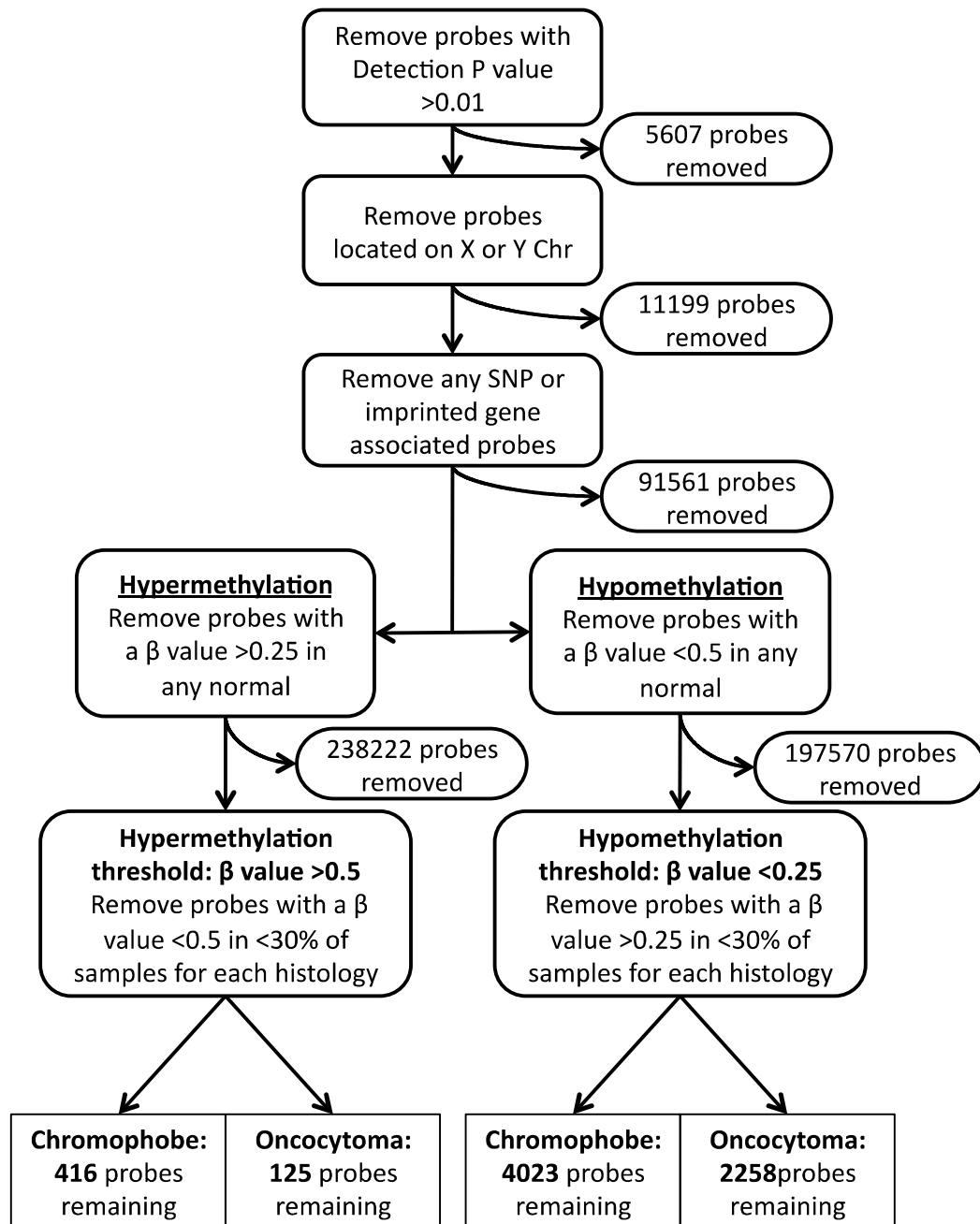
Figure 3-9: Flow chart depicting the steps undertaken to process 450K data

This was followed by the removal of all probes located over known SNPs and known imprinted genes, which may also skew and misrepresent the methylation profile (Chen et al., 2013). A total of 91561 such probes were removed.

Following removal of probes associated with SNPs or imprinted genes, the data processing diverged depending upon whether cancer specific hyper- or hypomethylation was to be identified.

3.2.1.3.1 *Cancer specific hypermethylation*

As previously described hypermethylation was defined as a β value of >0.5, however to identify cancer specific hypermethylation, the difference in β values between tumour samples and normal kidney samples should be at least a 2 fold. Therefore, any probe that presented with a β value >0.25 in any of the 4 normal kidney samples was omitted. This resulted in the removal of 238222 probes. The remaining probes were then utilised for further analysis into cancer specific hypermethylation (β value >0.5). The criteria of removing probes in which the normal samples β value of >0.25 was originally utilised in several HumanMethylation 27 Beadchip assay as the defining point between methylated and unmethylated loci (Hill et al., 2011; Morris et al., 2011).

3.2.1.3.1 *Cancer specific hypomethylation*

In order to study cancer specific hypomethylation, the opposite criteria was applied. Any probe presenting with a β value <0.5 in any on the 4 normal kidney samples was removed. A total of 197570 probes were removed. All remaining probes were then ready to be analysed for cancer specific hypomethylation. To conform with the 2 fold criteria

applied for hypermethylated probes (section 3.2.1.3.1) hypomethylation was determined to be any probe with a β value <0.25 within the cancer samples.

3.2.1.4 *Global methylation profile*

3.2.1.4.1 *Overview of methylation*

To ascertain the methylation profile of the oncocytoma and chRCC samples compared to the frequently studied pRCC and ccRCC histologies, we first identified the number of cancer specific probes presenting with hypermethylation >0.5 or hypomethylation <0.25 in at least one sample between the 4 histologies. A small randomly selected cohort of 30 papillary and 30 ccRCC were downloaded from the TCGA (sample IDs listed in Table 7-4 and Table 7-5) for this mini study. The number of probes with at least one sample meeting our criteria for hyper or hypomethylation were counted (Figure 3-10). It was observed that the more malignant tumours pRCC and ccRCC, presented with a higher proportion of hypermethylated probes (83.87% pRCC; 81.38% ccRCC cancer specific probes) than hypomethylated probes (16.13% pRCC; 18.62% ccRCC cancer specific probes). When analysing our chRCC and oncocytoma data, the reverse trend was observed with hypomethylated probes (71.64% chRCC; 81.40% oncocytoma cancer specific probes) being more prevalent that hypermethylated probes (28.36% chRCC; 18.60% oncocytoma cancer specific probes).
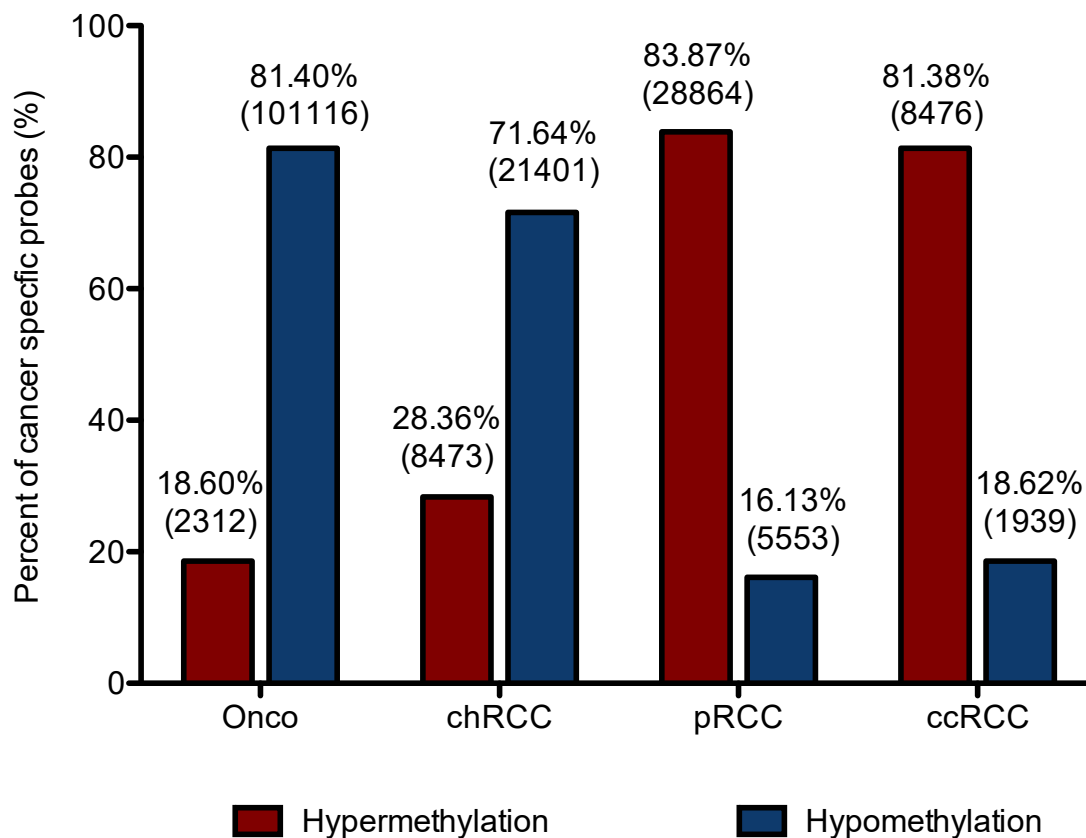
Figure 3-10: Cancer specific methylation profile for RCC subtypes
Methylation profiles were characterised as all cancer specific probes presenting with either hypermethylated (β value >0.5) or hypermethylated (β value <0.25) in at least 1 sample.

Cancer specific hypomethylation was observed to be more prevalent in the less aggressive subtypes of RCC investigated, chRCC (71.64% of chRCC cancer specific methylated probes were hypomethylated) and the benign renal oncocytoma (81.40% of oncocytoma cancer specific methylated probes were hypomethylated). Cancer specific hypermethylation on the other hand was observed to be more prevalent in the more malignant RCC subtypes pRCC (83.87% of cancer specific methylated probes were hypermethylated) or ccRCC (81.38% of ccRCC cancer specific methylated probes were hypermethylated). Chromophobe (chRCC) n=20 and Oncocytoma (Onco) n=21 were our samples, while papillary (pRCC) n=30 and clear cell (ccRCC) n=30 were randomly selected and downloaded from the TCGA.

3.2.1.4.2 *Clustering analysis*

To analyse the overall methylation profile of the chRCC and oncocytomas subtypes, we conducted, supervised hierarchical clustering using Euclidean distance complete linkage for the 500 most variable (probes with the greatest standard deviation between samples) cancer specific hypermethylated loci (Figure 3-11a) and hypomethylated loci (Figure 3-11b) was conducted.

Within the hypermethylation cluster (Figure 3-11a), two major cluster groups were identified. Cluster I was the largest cluster (n=34) and contained samples associated with lower levels or an absence of methylation. Within this cluster, a further 3 sub-clusters were identified, in which samples appeared to further segregate into clusters of the same histology. For example, Cluster I-I presented with a total of 9 samples; predominantly oncocytoma (7:2 oncocytoma: chRCC). Cluster I-II presented with 8 samples, all of which were oncocytoma, while Cluster I-III presented with 16 samples, predominantly chRCC; subdivided into 6:10 oncocytoma: chRCC, with 1 chRCC sample not falling into any of the three sub-clusters. The second major cluster (II) consisted of a total of 7 chRCC samples, 6 of which were sub-clustered and one sample excluded from the cluster, which presented with the highest levels of widespread methylation, reminiscent of a CIMP.

For the clustering analysis of hypomethylated probes (Figure 3-11b), segregation into the two distinct histologies was once again observed, with a total of 3 major clusters identified. Cluster I contained samples with the lowest methylation levels, and consisted of 2 subgroups with a total of 14 oncocytoma samples between them, and 1 chRCC not

141

A: Cancer specific hypermethylated probes   **B**: Cancer specific hypomethylated probes

β value

	1.0
	0.8
	0.6
	0.4
	0.2
	0.0

Histology
Gender
Age

		Key						
		Hypermethylation				Hypomethylation		
	Cluster:	1.1	1.2	1.3	2	1	2	3
Histology	Oncocytoma	n = 7	n = 8	n = 6	n = 0	n = 14	n = 2	n = 5
	Chromophobe	n =2	n = 0	n = 11	n = 7	n = 1	n = 15	n = 4
Gender	Male	n = 2	n = 4	n = 6	n = 2	n = 1	n = 6	n = 3
	Female	n = 7	n =3	n = 7	n = 2	n = 9	n = 6	n = 4
	No Data	n = 0	n = 1	n= 4	n = 3	n = 1	n = 5	n = 2
Age (years)	30-49	n = 2	n = 0	n = 5	n = 0	n = 2	n = 2	n = 3
	50-69	n = 4	n = 5	n = 5	n = 2	n = 9	n = 4	n = 3
	70-89	n = 3	n = 1	n = 3	n = 2	n = 3	n = 6	n = 1
	No Data	n = 0	n = 1	n= 4	n = 3	n = 1	n = 5	n = 2

I.I   I.II   I.III

I   II

I   II   III

Figure 3-11: Supervised hierarchical clustering using Euclidean distance with complete linkage

Clustering of the 500 most variable cancer-specific hypermethylated (A) and hypomethylated (B) loci. Clusters are denoted below the heatmap. Low β values (0) are represented as green and high β values (1) are red. Sample information is displayed below the cluster tree with chromophobe (orange) and oncocytoma (blue). Gender: female (black) and male (grey), and patient age 30–49 y (light purple); 50–69 y, (mid purple); 70–89 y, (dark purple). Crosses denote missing data.

142

included in either sub-cluster. The second and third clusters showed less hypomethylated probes than cluster I and predominantly contained chRCC samples with Cluster II containing 2 oncocytoma and 15 chRCC. The final cluster, Cluster 3 identified samples with the least hypomethylated probes, and contained 5 oncocytoma samples and 4 chRCC samples.

Both cluster analysis criteria undertaken showed no segregation or clustering within the clinical data available (age or gender) (Table 2-1). Furthermore, from this clustering analysis it was identified that the methylation profiles of renal oncocytoma and chRCC are distinct and can be separated, with chRCC presenting with more hypermethylated probes, while oncocytomas showed a pattern for more hypomethylated probes.

3.2.1.4.3 *Frequent methylation profile*

Within this study, frequent differential methylation was determined as all cancer specific probes presenting with either hyper- or hypomethylated in >30% of samples for each RCC histology, when compared to the normal kidney. This criteria identified the global differential methylation profile for frequently aberrant methylation (hyper and hypo) for each histology, compared to normal kidney DNA . 1.2% (4,439) loci were identified that met this criteria for chRCC samples while for oncocytoma, the frequency was lower, with 0.6% (2,383) of loci identified (Figure 3-12a). Within the frequently differential methylated cancer specific probes identified for chRCC, 416 probes were frequently hypermethylated (Table 7-7) and 4023 probes were frequently hypomethylated in chRCC . For oncocytoma 125 probes were frequently hypermethylated (Table 7-8) and 2258 probes were identified to be frequently hypomethylated (Figure 3-12a).
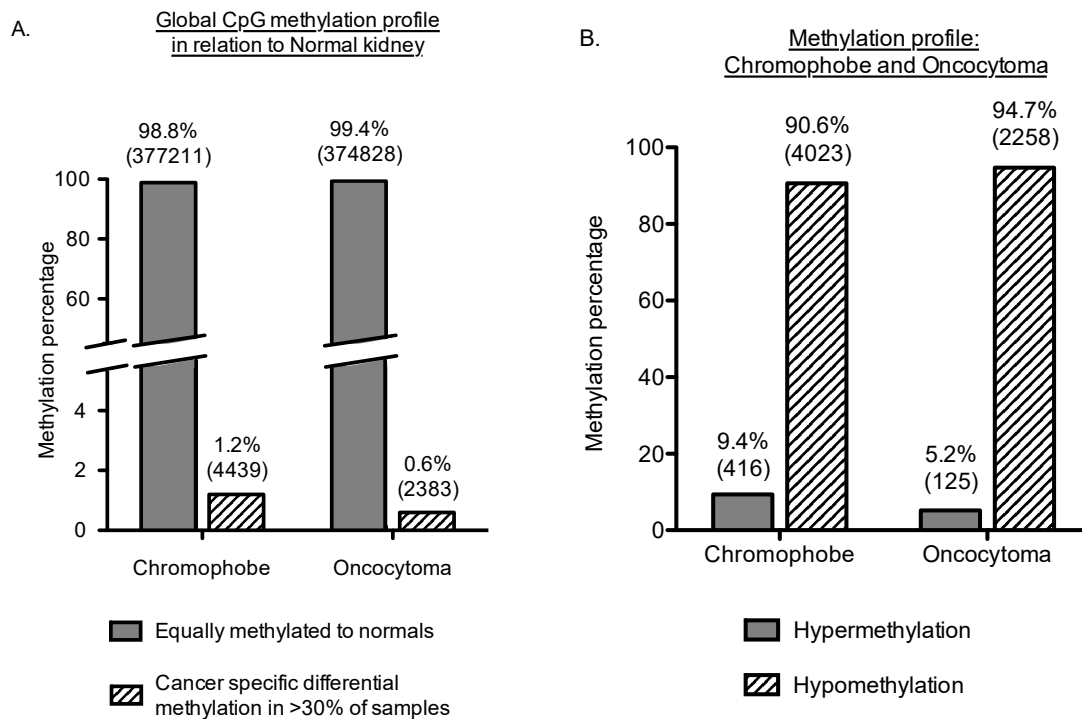
Figure 3-12: Methylation profile of chRCC and oncocytoma
A) The global methylation profile of differentially methylated cancer specific probes in relation to normal kidney. Defined as probes that are either frequently hypermethylated (β value >0.5) or hypomethylated β value (β value <0.25) in >30% of samples. 1.2% of all probes were identified to be frequently differentially methylated cancer specific loci for chRCC and 0.6% of probes for oncocytoma. **B)** Breakdown of differentially methylated cancer specific probes, showing the proportions that were identified as hypermethylated (9.4% chRCC and 5.2% oncocytoma) and those that were hypomethylated (90.6% chRCC and 94.7% oncocytoma)

The total frequent cancer specific hyper- or hypomethylated probes was identified to only account for a small proportion of all CpG probes interrogated. A total of 1.2% of CpG probes were identified as frequently methylated and cancer specific (combined hyper and hypo) in chRCC samples while for in oncocytomas only 0.6% of probes were identified as frequently methylated and cancer specific (combined hyper and hypo). However, when comparing frequently hypomethylated (>30% samples) cancer specific probes to the normal, both cancer histologies identified a large proportion of probes to be cancer specific, with 4,023 probes (90.6%) frequently hypomethylated and cancer specific in chRCC, and 2,258 probe (94.7%) identified for oncocytoma (Figure 3-12 and Figure 3-10). Frequent hypermethylation was identified to account for 9.4% of all chRCC frequently methylated cancer specific probes (416 loci, mapped to 204 genes; Table 7-7), while for oncocytoma 5.2% of frequently methylated cancer specific probes were hypermethylated (125 loci mapped to 70 refseq genes; Table 7-8).

To investigate the possible relevance of the hyper- and hypomethylation profile, the genomic distribution of identified probes in relation to Refseq genes was investigated (Figure 3-13a). Little variation in genomic distribution was identified between the two histologies, nor was much variation identified when analysing the percentage of probes located within the promoter region (defined as probes within the TSS1500, TSS200, 5'UTR or 1st exon), 3'UTR or gene body of Refseq genes when comparing between hyper- and hypomethylation.

However, variation was identified when investigating probe location in relation to CpG islands (Figure 3-13b).
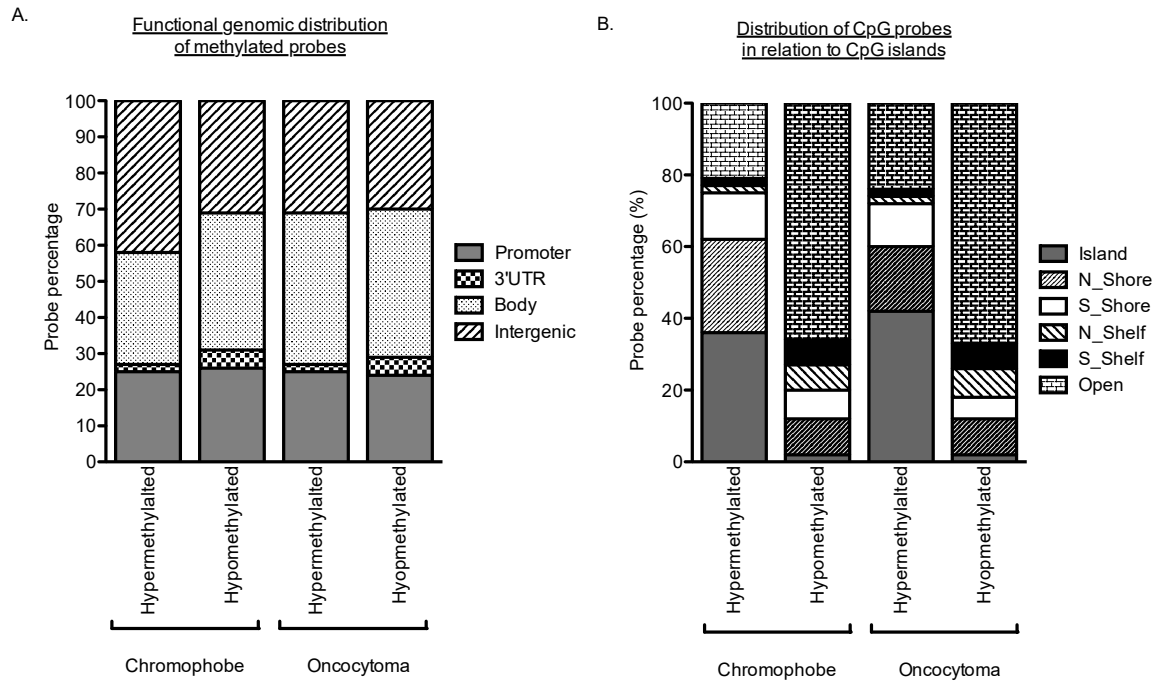
Figure 3-13: Genomic distribution of cancer specific hyper- and hypo-methylated loci

A) Genomic distribution of cancer-specific, hyper- and hypo-methylated CpG loci in relation known genes. The promoter region indicates loci residing within the 1st exon, 5'UTR, TSS200 and TSS1500. **B)** Genomic distribution of cancer-specific hyper- and hypo-methylated CpG loci in relation to CpG features and density. The majority of hypermethylated loci map to CG rich regions (predominately CpG islands, shores and shelves 79.1% for chromophobe samples and 76.0% for oncocytoma samples). Hypomethylated cancer-specific probes predominately map to lone CpG loci in open sea regions (65.7% for chromophobe samples and 67.3% for oncocytoma samples).

With hypomethylated probes showing a distinct trend to be located to lone isolated loci within the 'open sea' with 65.7% hypomethylated probes in chRCC and 67.3% hypomethylated probes in oncocytoma. Meanwhile hypermethylated probes were identified to predominantly be located in GC rich regions, associated with CpG islands and the surround areas (shores and shelves), with a total of 79.1% chRCC hypermethylated probes and 76.0% of oncocytoma hypermethylated probes located in these three regions.

3.2.1.5 *Analysis of hypermethylated probes*

Initial investigation of hypermethylated loci was focused on characterisation of the genes identified, and identification of interactions, networks and pathways that may be disrupted due the presence of abnormal methylation. To investigate this, IPA analysis was conducted on the full, frequently hypermethylated cancer specific gene list for each histology.

For the hypermethylated chRCC genes, the top three IPA identified networks included; embryonic development, tissue development and tissue morphology (13 associated genes), cell signalling, molecular transport and vitamin and mineral metabolism (12 associated genes) and cell death and survival, cell cycle and cancer (11 associated genes) (Figure 3-14, Table 7-7).

For the hypermethylated oncocytoma genes, IPA analysis identified the following key networks: cell signalling and molecular transport (10 associated genes), cell morphology, cellular assembly and organization, cellular function and maintenance,

Ingenuity associated functional networks of genes hypermethylated in >30% Chromophobe samples
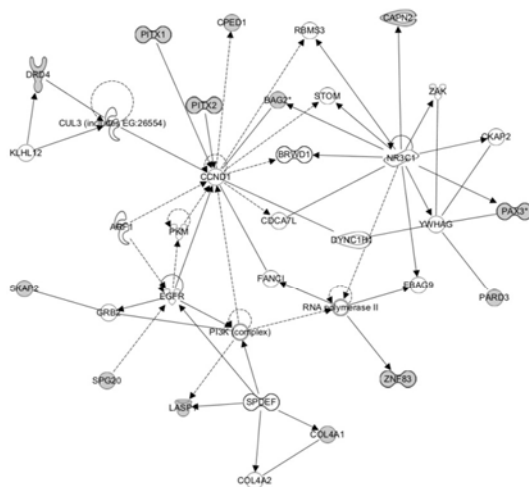
Network 1: 13 genes

Top functions: Embryonic development, Tissue Development and Tissue Morphology
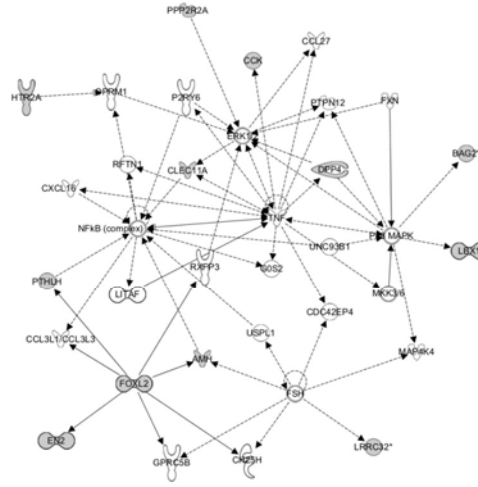
Network 2: 12 genes

Top functions: Cell signaling, Molecular transport, Vitamin and mineral metabolism.
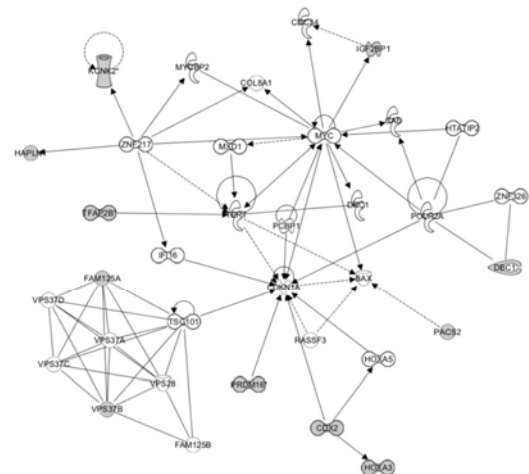
Network 3: 11 genes

Top functions: Cell death and survival, cell cycle and cancer.
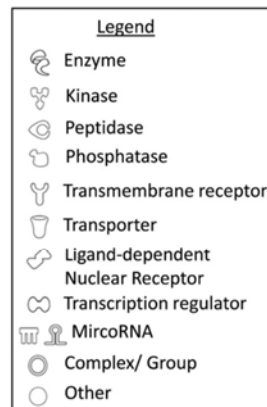
Figure 3-14: IPA analysis of frequently hypermethylated genes in chRCC
The top three networks identified by IPA functional network analysis of genes hypermethylated in > 30% of chromophobe RCC samples. Frequently hypermethylated genes are shown in grey and connecting genes in white. Solid arrows show direct interaction; dashed arrows represent indirect interactions and solid joining lines show protein binding.

Ingenuity associated functional networks of genes hypermethylated in >30% Oncocytoma
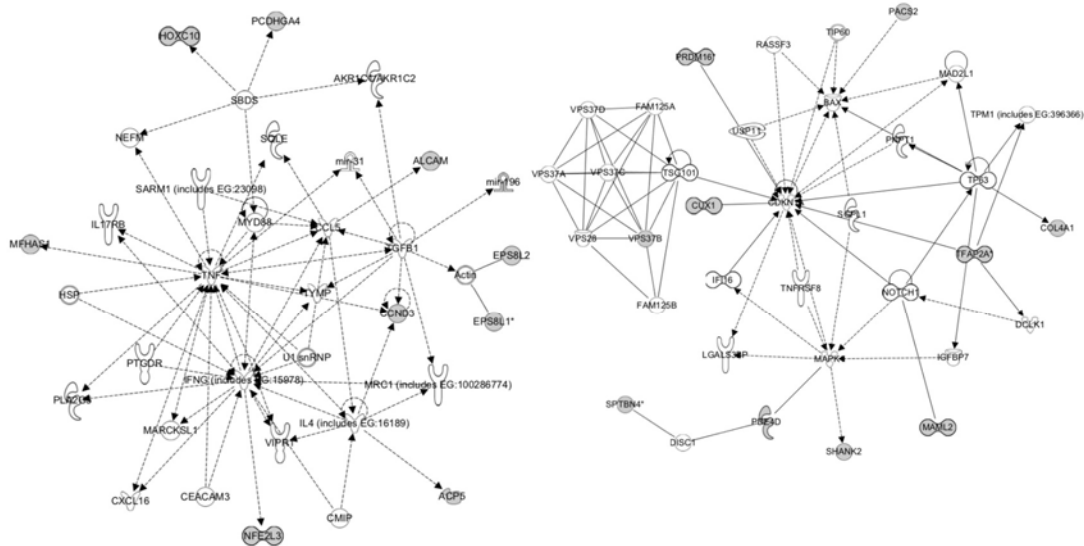
Network 1: 10 genes

Top functions: Cell signaling and molecular transport.

Network 2: 10 genes

Top functions: Cell morphology, cellular assembly and organization, cellular function and maintenance.
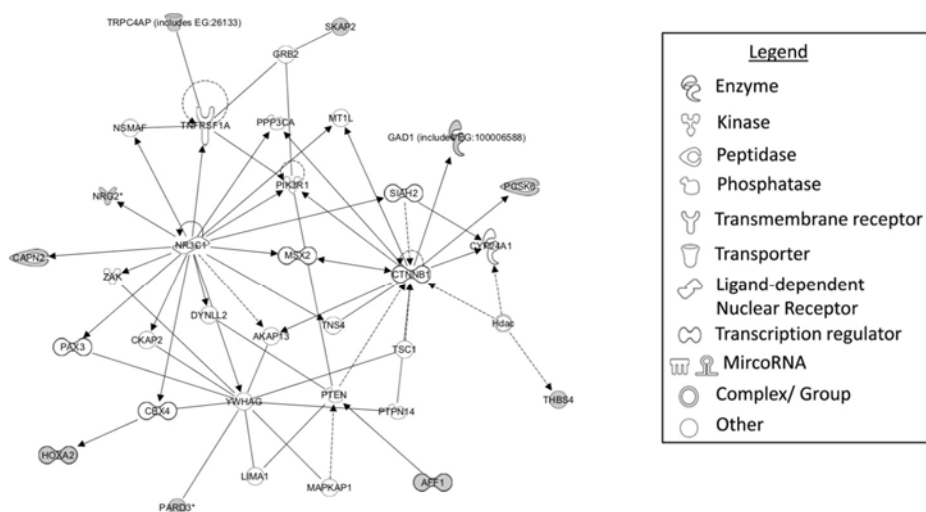
Network 3: 10 genes

Top functions: Cancer, cellular growth and proliferation, renal and urological disease.

Legend
- Enzyme
- Kinase
- Peptidase
- Phosphatase
- Transmembrane receptor
- Transporter
- Ligand-dependent Nuclear Receptor
- Transcription regulator
- MircoRNA
- Complex/ Group
- Other

Figure 3-15: IPA analysis of frequently hypermethylated genes in oncocytoma
The top three networks identified by IPA functional network analysis of genes hypermethylated in > 30% of chromophobe RCC samples. Frequently hypermethylated genes are shown in grey and connecting genes in white. Solid arrows show direct interaction; dashed arrows represent indirect interactions and solid joining lines show protein binding.

149

(10 associated genes) and cancer, cellular growth and proliferation and renal and urological disease (10 associated genes) (Figure 3-15, Table 7-8).

When comparing the frequently methylated cancer specific probes, between the two histologies we investigated, 48 genes (53 CG loci) were identified to be present in both the oncocytoma and chRCC data sets, which accounted for 17.5% of all gene identified as methylated. Of all the frequently hypermethylated genes identified, only 7 (*PCDH17, ASCL2, NKX6-2, HOXA9, PITX2, TLX3* and *ZNF177*) had previously been identified as methylated in other cancer studies (Arai et al., 2012; Costa et al., 2011; McRonald et al., 2009; Ricketts et al., 2012).

From the IPA analysis, it was shown that multiple networks may be affected, however the association of genes with cell survival, growth, cancer and other renal abnormalities suggests that the identified genes may be involved with the promotion or at least increased susceptibility to tumourigenesis.

3.2.1.5.1 *Identification of differentially hypermethylated genes*

The next focus of the study was to identify differentially hypermethylated genes that may distinguish chRCC histology from oncocytoma histology at the epigenetic level.

Differential hypermethylation was defined as CpG loci hypermethylated (β value > 0.5) in more than 30% of samples in one histology, and a β value < 0.3 in 90% of the samples of the other. Statistical significance was defined by Student's t-test analysis with FDR correction.

This approach identified 30 genes, which were significantly differentially hypermethylated, 28 of which were chromophobe-specific genes and 2 of which were oncocytoma-specific genes (ALCAM cg05645404 and TRPC4AP cg01154966; FDR P<0.05 (Table 3-1).

All except *SOX2OT* cg24513480 (North shelf) resided in the CpG rich island or shores. Genomics locations for these differentially hypermethylated genes varied with 24 probes located within the genes promoter region, broken down as 4 in the TSS1500 region, 3 probes in the TSS200 region, 3 in the 5' UTR and 1 in the 1st exon, while 15 probes were located in the gene body and 4 in the 3' UTR. Furthermore, 12 of the probes were mapped to a CpG rich region associated with a gene promoter. Probes that resided within promoter regions and within GC rich regions were henceforth referred to as promoter CGIs probes and are considered highly interesting and likely to influence gene expression.

For all differentially hypermethylated genes, pathway and network analysis was conducted using DAVID, Panther and IPA analysis. DAVID and Panther analysis identified several key pathways associated with the hypermethylated genes, including the Wnt signalling pathway (*EN2*), the MAPK signalling pathway (*CACNG7*), the Hippo pathway (*NPHP4*), the TGFβ signalling pathway (*AMH*) and associations with highly regulated processes such as cell death and apoptosis (*SPG20, NKX6-2, PAX3, BAG2*), as well as other functions such as the cell cycle, cell migration and cell adhesion that are frequently reported to be deregulated in cancer.

Gene Symbol	Cancer	Target ID	Relation to Gene	CpG region	Meth in Onco (B value >0.5)	Meth in Chromo (B value >0.5)	Normal Kidney	FDR P Value	Chr
ALCAM	Oc	cg05645404	Body	N_Shore	12/21 (57.1%)	0/20 (0.0%)	0/4	4.2E-05	3
TFAP2B	Ch	cg05437823	3'UTR	Island	0/21 (0.0%)	8/20 (40.0%)	0/4	1.91E-03	6
TRPC4AP; TRPC4AP	Oc	cg01154966	TSS1500; TSS1500	Island	7/21 (33.3%)	0/20 (0.0%)	0/4	3.40E-03	20
HOXA9	Ch	cg03217995	Body	N_Shore	0/21 (0.0%)	8/20 (40.0%)	0/4	3.65E-03	7
DBC1	Ch	cg03625109	TSS1500	Island	0/21 (0.0%)	7/20 (35.0%)	0/4	3.79E-03	9
CACNG7	Ch	cg21477176	3'UTR	S_Shore	1/21 (4.8%)	6/20 (30.%)	0/4	4.28E-03	19
NKX6-2	Ch	cg11174855	3'UTR	Island	1/21 (4.8%)	7/20 (35.0%)	0/4	4.33E-03	10
NPHP4	Ch	cg20383686	TSS200	Island	0/21 (0.0%)	6/20 (30.%)	0/4	4.80E-03	1
AMH	Ch	cg05345154	Body	Island	0/21 (0.0%)	6/20 (30.%)	0/4	6.46E-03	19
DPP4	Ch	cg19350270	Body	N_Shore	0/21 (0.0%)	7/20 (35.0%)	0/4	6.58E-03	2
SOX2OT	Ch	cg24513480	Body	N_Shelf	1/21 (4.8%)	8/20 (40.0%)	0/4	7.24E-03	3
EN2	Ch	cg12034383	TSS1500	Island	0/21 (0.0%)	6/20 (30.%)	0/4	7.26E-03	7
SPG20; SPG20; SPG20; SPG20	Ch	cg10558887	5'UTR; 5'UTR; 5'UTR; 5'UTR	N_Shore	0/21 (0.0%)	7/20 (35.0%)	0/4	7.49E-03	13
RALYL; RALYL; RALYL; RALYL; RALYL	Ch	cg22403811	5'UTR; TSS1500; 5'UTR; 5'UTR; 1stExon	N_Shore	0/21 (0.0%)	6/20 (30.%)	0/4	0.0096	8
JSRP1	Ch	cg04887494	Body	Island	0/21 (0.0%)	6/20 (30.%)	0/4	0.0098	19
PAX3; PAX3; PAX3; PAX3; PAX3; PAX3; PAX3; PAX3; CCDC140	Ch	cg13767755	Body; Body; Body; Body; Body; Body; Body; Body; TSS200	N_Shore	0/21 (0.0%)	6/20 (30.%)	0/4	0.0184	2
PITX1	Ch	cg00396667	3'UTR	Island	1/21 (4.8%)	6/20 (30.%)	0/4	0.0188	5
SH3PXD2A	Ch	cg18735015	Body	Island	1/21 (4.8%)	6/20 (30.%)	0/4	0.0213	10
SIX2	Ch	cg24887265	Body	N_Shore	0/21 (0.0%)	6/20 (30.%)	0/4	0.0219	2
TOX2; TOX2; TOX2	Ch	cg10900455	5'UTR; Body; 5'UTR	Island	1/21 (4.8%)	6/20 (30.%)	0/4	0.0224	20
HAPLN1	Ch	cg12199221	TSS200	N_Shore	0/21 (0.0%)	6/20 (30.%)	0/4	0.0226	5
IRX6	Ch	cg01064265	Body	Island	1/21 (4.8%)	7/20 (35.0%)	0/4	0.0230	16
NKAPL	Ch	cg17384889	TSS200	Island	0/21 (0.0%)	6/20 (30.%)	0/4	0.0236	6
LBX1	Ch	cg03053579	Body	Island	0/21 (0.0%)	6/20 (30.%)	0/4	0.0237	10
BAG2	Ch	cg10230427	Body	S_Shore	1/21 (4.8%)	7/20 (35.0%)	0/4	0.0242	6
MKX	Ch	cg26298409	Body	Island	0/21 (0.0%)	6/20 (30.%)	0/4	0.0262	10
SLITRK1	Ch	cg16727923	1stExon	Island	1/21 (4.8%)	7/20 (35.0%)	0/4	0.0376	13
BAG2	Ch	cg27164797	Body	S_Shore	0/21 (0.0%)	6/20 (30.%)	0/4	0.0401	6
KRT27	Ch	cg02399249	Body	Island	0/21 (0.0%)	6/20 (30.%)	0/4	0.0421	17
HHEX	Ch	cg09721427	TSS1500	Island	0/21 (0.0%)	6/20 (30.%)	0/4	0.0476	10

Table 3-1: Differentially hypermethylated genes between chRCC and oncocytoma
Table of the differentially hypermethylated probes, defined as β value >0.5 in >30% samples of one histology and β value <0.3 in 90% of the second histology. Multiple gene names represent the different transcripts to which a CpG locus maps. The CpG location in 'relation to gene' is then displayed for each transcript. CH, chromophobe; Oc, oncocytoma; Chr, Chromosome

Moreover, IPA analysis of the complete gene list, identified genes to be associated with networks involved in connective tissue development and function, embryonic and organ development, cancer, reproductive system disease and cellular development (Figure 3-16). The next stage of the study was to investigate whether any of the differentially hypermethylated loci identified were specific for either chRCC or oncocytoma, when compared to TCGA data for the RCC subtypes ccRCC and pRCC. To achieve, this the 28 differentially hypermethylated chRCC loci previously identified (section 3.2.1.5.1), were now compared to 450K array methylation data for papillary RCC (n=81) and ccRCC (n=100) samples downloaded from the TCGA (section 2.8.2). One-way Anova with Games-Howell post hoc test was performed. From this analysis, 3 genes (*SPG20*, *NPHP4* and *TFAP2B*) were identified to be significantly hypermethylated in chRCC compared to oncocytoma, ccRCC and pRCC. *SPG20* and *NPHP4* loci were shown to reside within the CpG dense regions for the North shore and island and to be associated with the promoter regions of the associated gene, while *TFAP2B cg05437823* was located in the CpG island mapped to the 3'UTR of the gene.
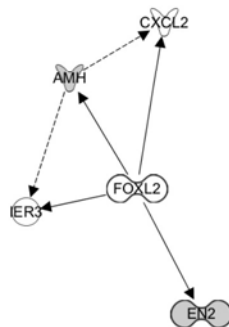
Using the same method described above the 2 genes identified to be differentially hyper-methylated in oncocytoma were also compared to ccRCC and pRCC. This processes identified *ALCAM* cg05645404 to be the only locus significantly differentially hypermethylated in oncocytoma when compared to the three other renal cancer histologies (Figure 3-17).

Figure 3-16: IPA analysis of differentially hypermethylated genes in between chRCC and oncocytoma

The top two networks identified by IPA functional network analysis of differentially methylated genes between chRCC and oncocytoma. Differential hypermethylation was defined as defined as β value >0.5 in >30% samples of one histology and β value <0.3 in 90% of the second histology hypermethylated in > 30%. Frequently hypermethylated genes are shown in grey and connecting genes in white. Solid arrows show direct interaction; dashed arrows represent indirect interactions and solid joining lines show protein binding.
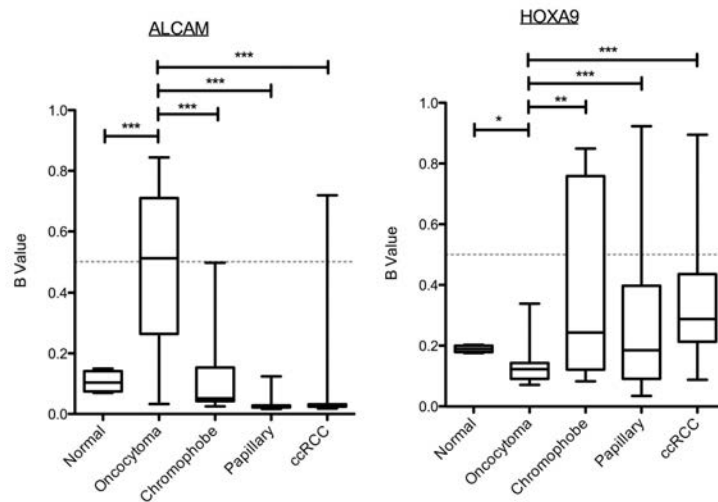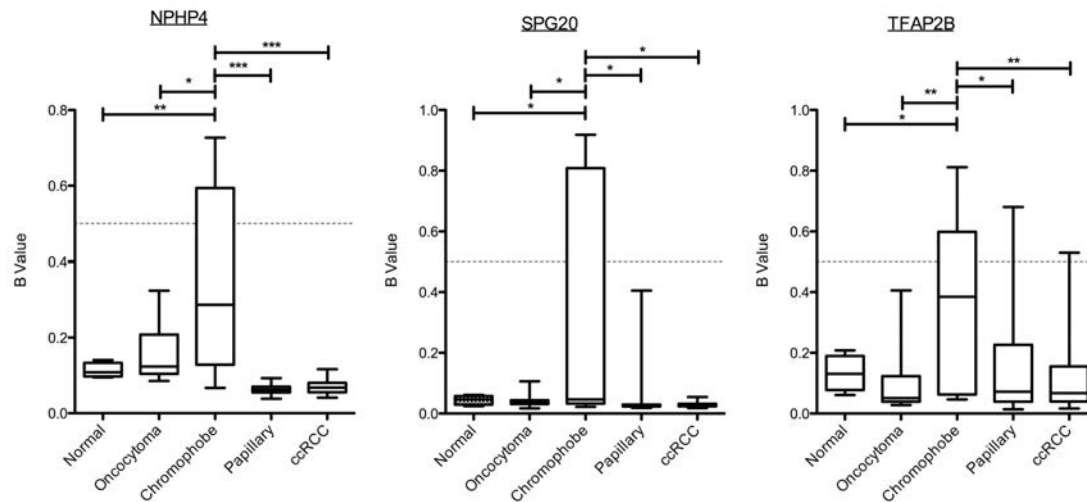
Figure 3-17: Histology specific differential methylation
Differentially hypermethylated genes identified between chRCC and oncocytoma were compared to the TCGA data for ccRCC and pRCC. Significance was calculated using one-way ANOVA with Games-Howell post hoc test. β values are displayed as box plots with whiskers showing the median, maximum and minimum value for each histology **A)** Oncocytoma: One gene was significantly hypermethylated in oncocytoma *ALCAM*, while *HOXA9* showed significant reduction in hypermethylation of oncocytoma. **B)** chRCC: Three genes were identified to be significantly hypermethylated in chRCC. CpG loci investigated, *ALCAM:* cg05645404, *HOXA9*: cg03217995, *NPHP4*: cg20383686, *SPG20:* cg10558887 and *TFAP2B:* cg05437823. p values: * < 0.05, ** < 0.01, *** < 0.001

155

The *ALCAM* CpG locus mapped to the north shore and the refSeq gene body. However, for *HOXA9,* a gene originally investigated due to differential hypermethylation in chRCC. We observed instead that of presenting with chRCC specific methylation, *HOXA9* cg*03217995,* displayed as significantly hypermethylated in all investigated renal cancer histologies when compared to oncocytoma. In oncocytoma average β-value was across all samples was 0.13 and thus were deemed unmethylated

3.2.1.5.2 *Confirmation of differential methylation*

Due to the aforementioned association of *NPHP4* and *SPG20* with the Hippo pathway, cell death and apoptosis respectably, single colony bisulfite sequencing was conducted on 2 chRCC samples presenting with methylation and 2 oncocytoma samples lacking methylation. Bisulfite sequencing confirmed the differential methylation pattern identified by the 450K array (Figure 3-18, Figure 3-19).

To confirm frequent methylation in chRCC, a separate cohort of 22 Japanese chRCC were analysed for methylation of *SPG20* and *NPHP4* by CoBRA analysis. Here we identified that methylation was present at varying frequencies in 54.5% and 77.3% of chRCC samples for *SPG20* and *NPHP4* respectively (Figure 3-20). This is higher than the frequencies observed via the 450K array in which 35.0% of chRCC samples presented with *SPG20* methylation and 30.0% chRCC samples presented with methylation of *NPHP4*. This variation could be attributed to the technique employed to assess the methylation status, as the 450K is more sensitive identifying methylation at a individual CpG locus, rather than Cobra, which identifies the presence of methylation within a

region of interest (Xiong and Laird, 1997). Secondly variation in methylation frequency could be due to the second cohort being of different ethnic origin and exposed to a different environment, effecting levels of methylation (Busche et al., 2015; Heyn et al., 2013)
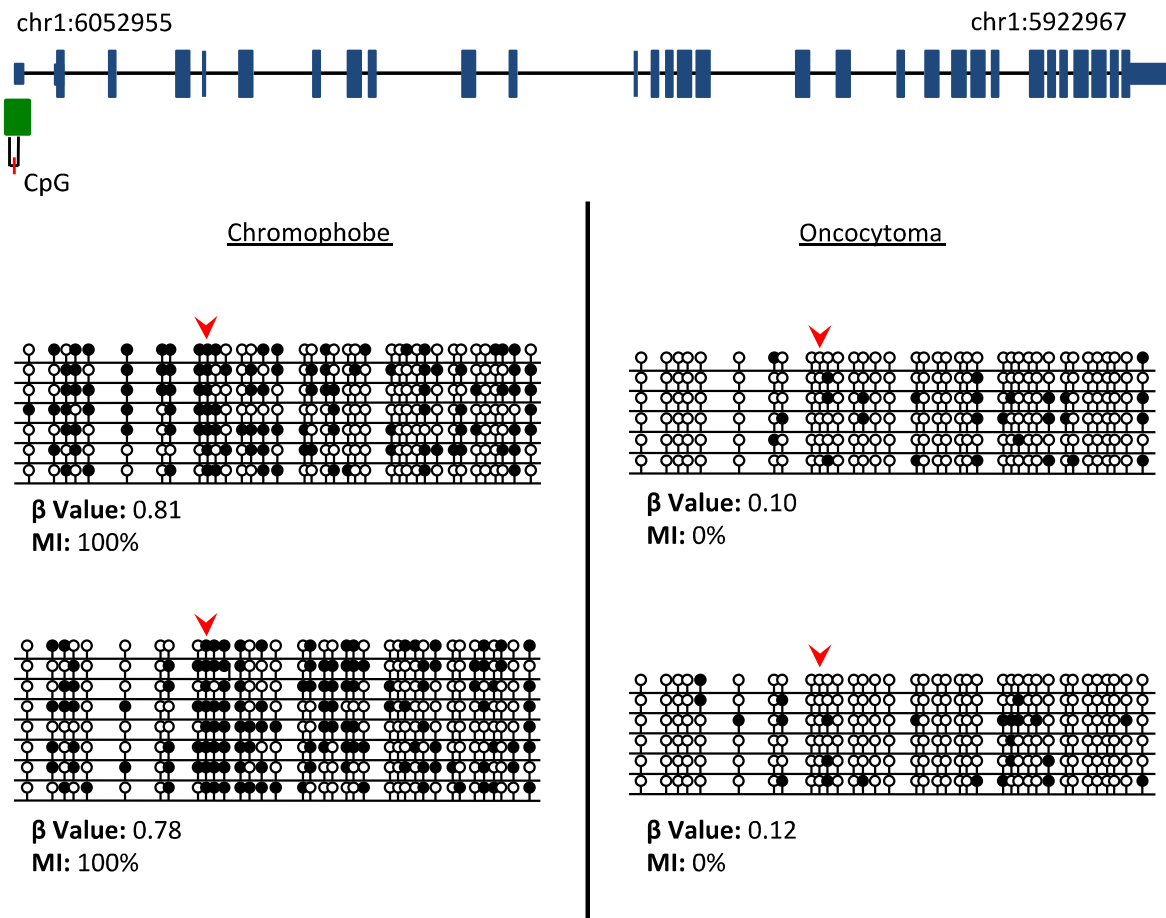
Figure 3-18: Bisulfite sequencing of *NPHP4* to confirm the methylation status of locus cg20383686

Methylation of *NPHP4* cg20383686 was confirmed by bisulfite sequencing four samples (2 chRCC and 2 oncocytoma). The *NPHP4* CpG locus maps to TSS200 and the CpG island. MI values confirmed the β values observed with hypermethylation of the locus (β value > 0.5) presenting with a high MI > 50% confirming that the locus is methylated (chRCC), while the unmethylated β value locus correlated with a low MI, low methylation (oncocytoma).

Diagram show schematic of the gene (blue) and CpG island (green) with region interrogated by bisulfite sequencing marked by the black box. The CpG locus is represented by the red mark on gene schematic and red arrows on the bisulfite clone sequencing readout. Each block of bisulfite sequencing is a samples, with each rows representing an individual allele, circles indicating location and methylation state of a CpG locus (black circle: methylated; white circle: unmethylated).

158

SPG20: NM_001142296   **CpG loci:** *CG*10558887

Ch13: 36921004   Ch13: 36875775

CpG

Chromophobe   Oncocytoma

β Value: 0.84
MI: 100%

β Value: 0.05
MI: 0%

β Value: 0.73
MI: 100%

β Value: 0.04
MI: 0%
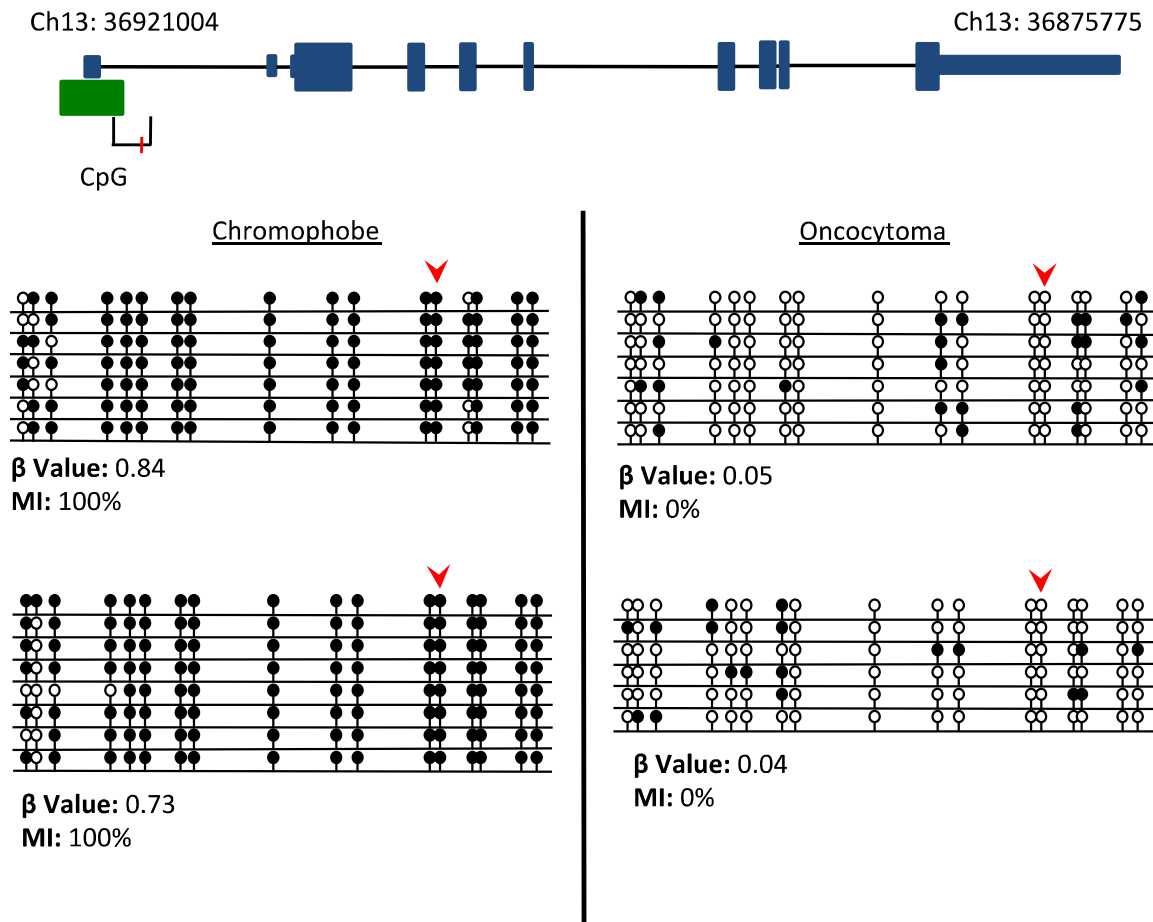
Figure 3-19: Bisulfite sequencing of *SPG20* to confirm the methylation status of locus cg10558887

Methylation of *SPG20* cg10558887 was confirmed by bisulfite sequencing four samples (2 chRCC and 2 oncocytoma). The *SPG20* CpG locus maps to the 5'UTR and the Nor. MI values confirmed the β values observed with hypermethylation of the locus (β value > 0.5) chRCC presenting with a high MI > 50% confirming that the locus is methylated, while the unmethylated β value locus correlated with a low MI, low methylation oncocytoma.

Diagram show schematic of the gene (blue) and CpG island (green) with region interrogated by bisulfite sequencing marked by the black box. The CpG locus is represented by the red mark on gene schematic and red arrows on the bisulfite clone sequencing readout. Each block of bisulfite sequencing is a samples, with each rows representing an individual allele, circles indicating location and methylation state of a CpG locus (black circle: methylated; white circle: unmethylated).
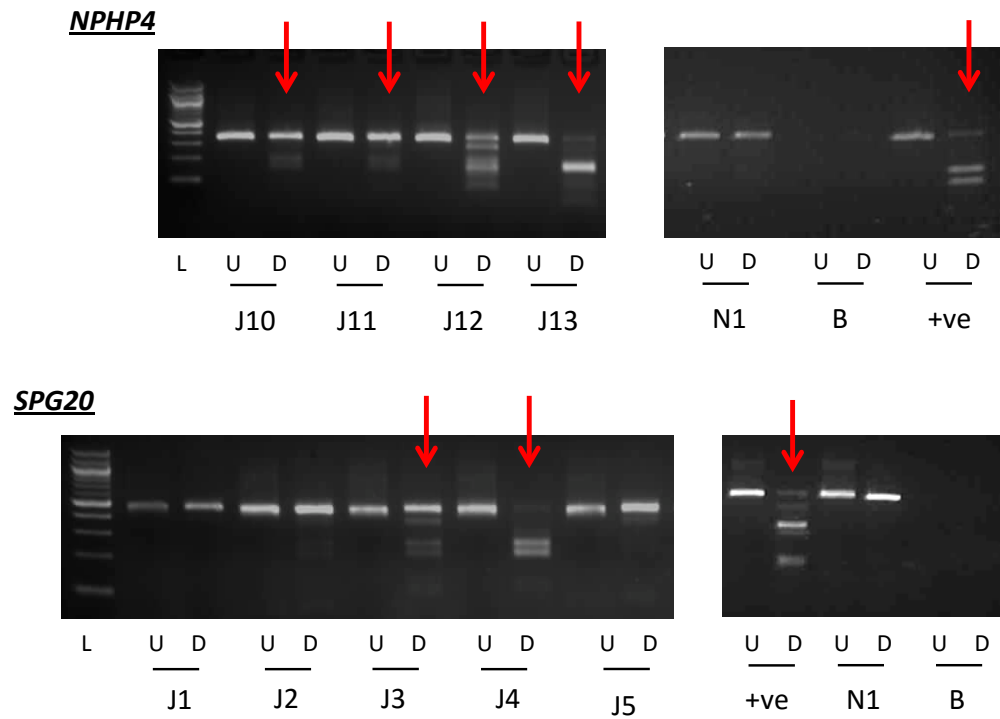
Figure 3-20: CoBRA analysis of *NPHP4* and *SPG20* methylation.
CoBRA analysis of *NPHP4* and *SPG20* methylation on an independent cohort of 22 primary Japanese chRCC. Methylation was observed in 54.5% and 77.3% of chRCC samples *NPHP4* and 54.5% of the samples tested for *SPG20*.

CoBRA PCR product was incubated with *Taqα1* and undigested PCR product (U) was run next to *Taqα1* digested product *(D)*, with digestion of the band is indicative of methylation (also marked by the red arrows).

3.2.1.6 *Analysis of hypomethylated probes*

Following on from the investigations undertaken in profiling the hypermethylation of chRCC and oncocytoma, a similar analysis was concurrently undertaken, investigating the hypomethylation profile between chRCC and oncocytoma histologies.

The study of hypomethylated loci initially focused on frequently differentially hypomethylated cancer specific CpG loci. Once again frequent hypomethylation was deemed to be loci presenting with β value <0.25 in > 30% of samples of each histology.

For the chRCC cohort, a total of 2,134 genes (4,023 loci) were deemed to be frequently hypomethylated. IPA analysis identified multiple networks, with the top three shown in Figure 3-21. Here IPA identified hypomethylated genes associated with networks involved in cell to cell signalling, hereditary disorders, neurological diseases, 35 post-translational modification genes, 33 tissue morphology genes, cancer and 33 gastrointestinal disease genes.

For the oncocytoma cohort, the same analysis identified 1,239 genes (2,258 loci) to be frequently hypomethylated. With further IPA analysis, predicting network interactions associated with hereditary disorders, neurological disease, skeletal and 33 muscular disorders genes, cancer, cellular development, 33 cellular function and maintenance genes, embryonic development, lymphoid tissue structure and development, and 30 organ development genes within the top three networks (Figure 3-22). Comparison of the two frequently hypomethylated probe lists identified 943 genes (1,543 loci) to be hypomethylated in both histologies, which accounted for 27.9% of all hypomethylated genes identified.

Ingenuity associated functional networks of genes hypomethylated in >30% Chromophobe samples

Network 1: 35 genes

Top functions: Cell to cell signaling, Hereditary disorders, Neurological disease.

Network 2: 33 genes

Top functions: post-translational modification, Hereditary disorder, Neurological disease.

Network 3: 33 genes

Top functions: Tissue morphology, Cancer, Gastrointestinal disease.

Legend

- Cytokine/ Growth Factor
- Enzyme
- Kinase
- Peptidase
- G-Protein coupled receptor
- Transmembrane receptor
- Ion Channel
- Transporter
- Transcription regulator
- MircoRNA
- Complex/ Group
- Other

Figure 3-21: IPA analysis of frequently hypomethylated genes in chRCC
The top three networks identified by IPA functional network analysis of genes hypermethylated in > 30% of chromophobe RCC samples. Frequently hypomethylated genes are shown in grey and connecting genes in white. Solid arrows show direct interaction; dashed arrows represent indirect interactions and solid joining lines show protein binding.

Network 1: 33 genes

Top functions: Hereditary disorder, Neurological disease, Skeletal and muscular disorders.

Network 2: 33 genes

Top functions: Cancer, Cellular development, Cellular Function and Maintenance.

Network 3: 30 genes
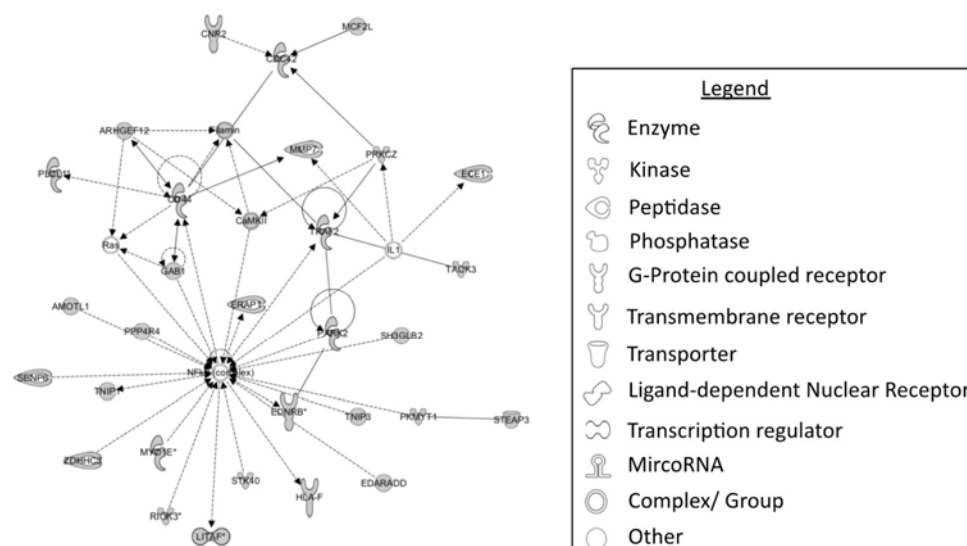
Top functions: Embryonic development, Lymphoid tissue structure and development, Organ development.

Figure 3-22: IPA analysis of frequently hypomethylated genes in oncocytoma
The top three networks identified by IPA functional network analysis of genes hypermethylated in > 30% of renal oncocytoma samples. Frequently hypomethylated genes are shown in grey and connecting genes in white. Solid arrows show direct interaction; dashed arrows represent indirect interactions and solid joining lines show protein binding.

3.2.1.6.1 *Identification of differentially hypomethylated genes*

Similar to the differential methylation analysis undertaken for hypermethylation, differentially hypomethylated genes were identified as loci with a β value < 0.25 in > 30% of the samples for one of the histologies, while the other histology presents with a β value > 0.45 in 100% of the samples. Student's t-test with FDR correction was conducted to identify significant differentiation. This analysis identified a total of 43 loci encompassing 41 genes; 5 genes in oncocytoma and 36 genes in chRCC samples (Table 3-2: Differentially hypomethylated). Of these 41 differentially hypomethylated genes, it was identified that 6 chRCC specific genes were located within promoter associated CpG shores and none within promoter CGIs.

IPA analysis of these 41 genes identified involvement with networks including cardiovascular system development and function, cell cycle, cell death and survival, cellular development, skeletal and muscular system development and function, and cellular growth and proliferation (Figure 3-23).

Gene Symbol	Cancer	Target ID	Relation to Gene	CpG region	Meth in Onco (B value <0.25)	Meth in Chromo (B value <0.25)	Normal Kidney	FDR P Value	Chr
ITGB5	Oc	cg00171092	Body	Open Sea	7/21 (33.3%)	0/20 (0.0%)	0/4	2.7E-08	3
GTF2IRD1; GTF2IRD1	Ch	cg20448594	Body; Body	Open Sea	0/21 (0.0%)	7/20 (35.0%)	0/4	2.91E-08	7
PABPC4L	Ch	cg15867829	3'UTR	Open Sea	0/21 (0.0%)	9//20 (45.0%)	0/4	1.20E-07	4
PRKCE	Ch	cg04035064	Body	Open Sea	0/21 (0.0%)	7/20 (35.0%)	0/4	2.97E-07	2
CTDSPL; CTDSPL	Oc	cg15428496	Body; Body	Open Sea	7/21 (33.3%)	0/20 (0.0%)	0/4	3.01E-07	3
PLCL1	Oc	cg02833180	Body	Open Sea	10/21 (47.6%)	0/20 (0.0%)	0/4	3.16E-07	2
CD44; CD44; CD44; CD44; CD44	Ch	cg25096745	Body; Body; Body; Body; Body	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	3.33E-07	11
TSPAN5	Oc	cg03234557	Body	Open Sea	9/21 (42.8%)	0/20 (0.0%)	0/4	8.95E-07	4
SCG3; SCG3	Ch	cg08884979	3'UTR; 3'UTR	Open Sea	0/21 (0.0%)	11/20 (55.0%)	0/4	4.81E-06	15
CSRP2BP; CSRP2BP	Ch	cg14329508	3'UTR; Body	Open Sea	0/21 (0.0%)	7/20 (35.0%)	0/4	1.54E-05	20
MIR141	Ch	cg19794481	TSS200	Open Sea	0/21 (0.0%)	6/20 (30.0%)	0/4	1.59E-05	12
SLC44A3; SLC44A3	Ch	cg22608160	Body; Body	Open Sea	0/21 (0.0%)	7/20 (35.0%)	0/4	3.39E-05	1
MICAL2	Oc	cg00286773	Body	Open Sea	7/21 (33.3%)	0/20 (0.0%)	0/4	3.57E-05	11
PHTF2; PHTF2	Ch	cg27457191	5'UTR; 5'UTR	S.Shore	0/21 (0.0%)	6/20 (30.%)	0/4	0.0000	7
RORA	Ch	cg03952578	Body	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	0.0000	15
ARID1B; ARID1B; ARID1B	Ch	cg04924555	Body; Body; Body	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	0.0001	6
LRIG1	Ch	cg11198596	Body	N.Shelf	0/21 (0.0%)	6/20 (30.%)	0/4	0.0001	3
COL18A1; COL18A1; COL18A1	Ch	cg25049844	3'UTR; 3'UTR; 3'UTR	S.Shore	0/21 (0.0%)	7/20 (35.0%)	0/4	0.0001	21
ITPRIP	Ch	cg16301004	5'UTR	Open Sea	0/21 (0.0%)	7/20 (35.0%)	0/4	0.0001	10
RAB3B	Ch	cg00714309	Body	Open Sea	0/21 (0.0%)	8/20 (40.0%)	0/4	0.0001	1
MIR141	Ch	cg02624246	Body	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	0.0001	12
GBE1	Ch	cg05514531	Body	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	0.0002	3
SLC27A3	Ch	cg15930240	Body	S.Shore	0/21 (0.0%)	6/20 (30.%)	0/4	0.0002	1
MKLN1; MKLN1	Ch	cg02776119	Body; Body	S.Shelf	0/21 (0.0%)	6/20 (30.%)	0/4	0.0002	7
C14orf169; HEATR4	Ch	cg02935494	TSS1500; Body	N.Shore	0/21 (0.0%)	9//20 (45.0%)	0/4	0.0002	14
VPS13B; VPS13B	Ch	cg18520777	Body; Body	Open Sea	0/21 (0.0%)	7/20 (35.0%)	0/4	0.0002	8
MYO3B; MYO3B; MYO3B	Ch	cg16066505	Body; Body; Body	Open Sea	0/21 (0.0%)	8/20 (40.0%)	0/4	0.0002	2
DSG1	Ch	cg07703790	TSS1500	Open Sea	0/21 (0.0%)	8/20 (40.0%)	0/4	0.0005	18

Gene Symbol	Cancer	Target ID	Relation to Gene	CpG region	Meth in Onco (B value <0.25)	Meth in Chromo (B value <0.25)	Normal Kidney	FDR P Value	Chr
RNPEP	Ch	cg16047663	3'UTR	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	0.0005	1
OR2B2	Ch	cg19173375	TSS200	Open Sea	0/21 (0.0%)	8/20 (40.0%)	0/4	5.88E-04	6
NEK11; NEK11; NEK11	Ch	cg06239593	Body; Body; Body	S.Shelf	0/21 (0.0%)	6/20 (30.%)	0/4	8.99E-04	3
NRP2; NRP2; NRP2; NRP2; NRP2	Ch	cg14157435	Body; Body; Body; Body; Body	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	1.21E-03	2
PLEKHG7; PLEKHG7	Ch	cg05916700	1stExon; 5'UTR	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	2.10E-03	12
CDKAL1	Ch	cg00682125	Body	Open Sea	0/21 (0.0%)	7/20 (35.0%)	0/4	3.11E-03	6
HOXC4; HOXC6; HOXC5	Ch	cg18040878	TSS1500; TSS1500; TSS1500	S.Shore	0/21 (0.0%)	6/20 (30.%)	0/4	4.61E-03	12
ACTN2	Ch	cg18381395	Body	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	5.00E-03	1
C6orf195	Ch	cg01285783	Body	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	5.48E-03	6
MGRN1; MGRN1; MGRN1; MGRN1	Ch	cg02074956	Body; Body; Body; Body	N.Shore	0/21 (0.0%)	6/20 (30.%)	0/4	9.04E-03	16
SVIL; SVIL	Ch	cg04678141	TSS200; 5'UTR	Open Sea	0/21 (0.0%)	6/20 (30.%)	0/4	1.49E-02	10
CAST; CAST; CAST; CAST; CAST	Ch	cg01741836	TSS1500; TSS1500; TSS1500; TSS1500; TSS1500	N.Shore	0/21 (0.0%)	6/20 (30.%)	0/4	1.69E-02	5
NDUFAF1	Ch	cg13897914	TSS1500	S.Shore	0/21 (0.0%)	6/20 (30.%)	0/4	2.34E-02	15
ACCN4; ACCN4	Ch	cg08614334	Body; Body	S.Shelf	0/21 (0.0%)	6/20 (30.%)	0/4	0.0275	2

Table 3-2: Differentially hypomethylated genes between chRCC and oncocytoma
Table of the differentially hypomethylated probes identified between chRCC and oncocytoma. Hypomethylation is defined as β value <0.25 in >30% samples of one histology and β value >0.4 in 100% of the second histology. Multiple gene names represent the different transcripts to which a CpG locus maps. The CpG location in 'relation to gene' is then displayed for each transcript. CH, chromophobe; Oc, oncocytoma; Chr, Chromosome

Ingenuity associated functional networks of significantly differentially hypomethylated genes hypomethylated in chromophobe and oncocytoma
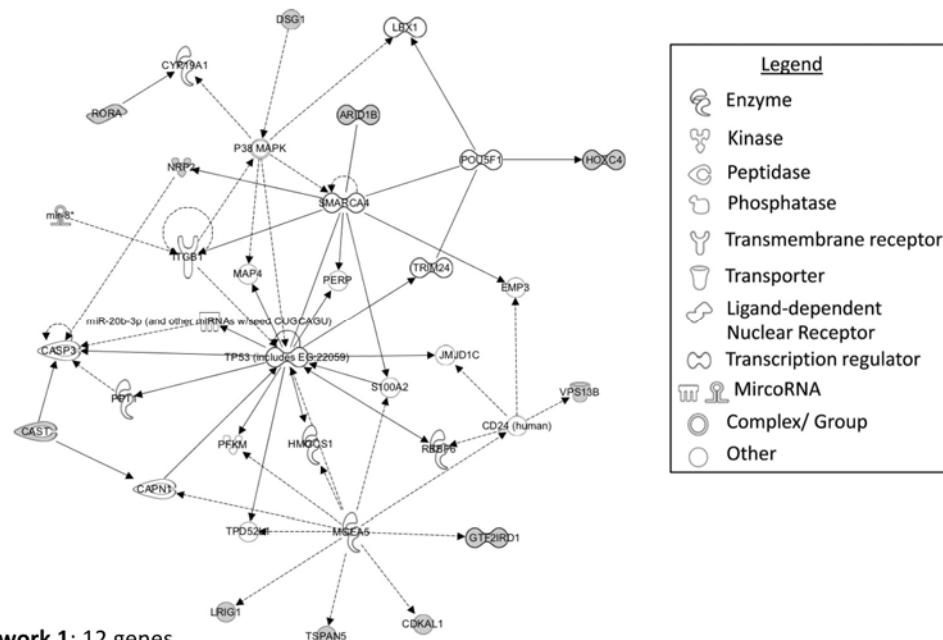
Network 1: 12 genes

Top functions: Cellular development, Skeletal and muscular system development and function, Cellular growth and proliferation.
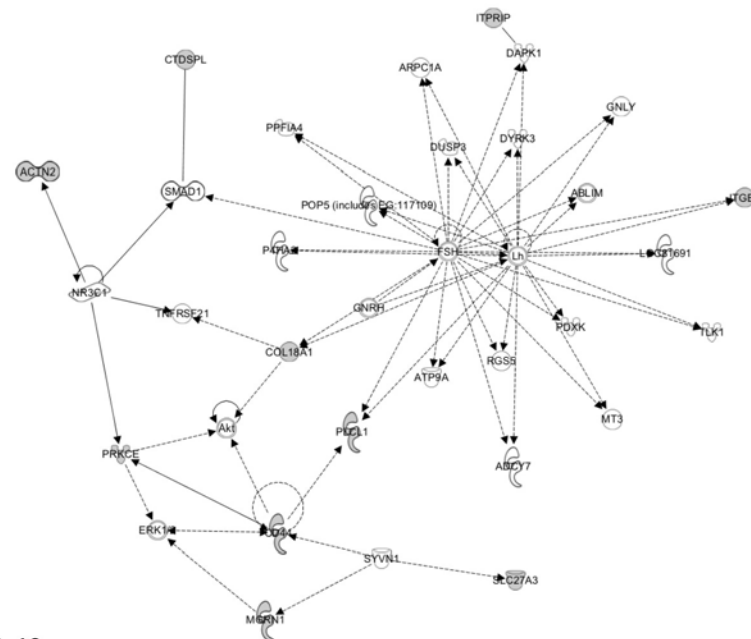
Network 2: 10 genes

Top functions: Cardiovascular system development and function, Cell cycle, Cell death and survival.

Figure 3-23: IPA analysis of differentially hypomethylated genes in between chRCC and oncocytoma (See next page for figure legend)

167

The top two networks identified by IPA functional network analysis of differentially methylated genes between chRCC and oncocytoma. Differential hypermethylation was defined as defined as β value <0.25 in >30% samples of one histology and β value >0.5 in 100% of the second histology hypermethylated in > 30%. Frequently hypomethylated genes are shown in grey and connecting genes in white. Solid arrows show direct interaction; dashed arrows represent indirect interactions and solid joining lines show protein binding.

3.3 <u>Discussion</u>

chRCC and oncocytoma are both believed to be derived from distal renal tubules and possess a similar morphology and appearance, that can be difficult to differentiate visually through imaging and biopsies due to only a fraction of the tumour being analysed, and the full cellular architecture not being represented (Ng et al., 2014; Pierorazio, 2014). Fast and accurate identification is vital to ensure appropriate treatment and management is undertaken. As a result, much focus has been invested on identifying biochemical and genetic differentiation between all renal carcinoma subtypes. Current methods include analysis of known chromosomal abnormalities frequently identified in the histologies, such as the loss of chromosome 1p and Y in oncocytoma or Chromosomes 1, 2, 6, 10 and 17 in chRCC (Yusenko et al., 2009), or use immunohistochemistry to differentiate between histologies. Examples of this include staining for CK7, which is strong in chRCC but weak in oncocytoma, and CD15, which is absent in chRCC but 70% oncocytomas stain positive, while staining for CD10 can distinguish between ccRCC, chRCC and oncocytoma being positive in ccRCC and negative in the other two histologies. Hale's or Modified Mowry's colloidal iron stain is reported to strain more strongly in chRCC than oncocytomas, while comparison of cadherin gene family expression can also distinguish between histologies, with kidney-specific cadherin almost exclusively being expressed in chRCC. However none of the cytogenetic or immunohistochemistry techniques are regarded as accurate or reliable at differentiating the two histologies (Ng et al., 2014).

3.3.1 Summary of findings

Within this study, we aimed to identify the methylation profiles of rare RCC subtypes using a whole genome wide high throughput methylation array, the Infinium HumanMethylation 450 BeadChip.

3.3.1.1 *Genome wide methylation analysis*

The methylation profiles of 21 oncocytoma and 20 chRCC were analysed using the Infinium HumanMethylation 450 BeadChip, where >480,000 CpG loci were interrogated for methylation status. The array was validated in house by single colony bisulfite sequencing of three candidate genes, while the reproducibility of the array was ascertained by the inclusion of a duplicate sample on the array. Following a manually developed pipeline to identify cancer specific probes, we identified that chRCC and oncocytoma exhibit similar patterns of global DNA methylation with the highest proportion of cancer specific probes presenting with hypomethylation (>70% cancer specific probes), while <30% of all cancer specific probes were identified to be hypermethylated. The benign oncocytoma presented with the higher percentage of hypomethylated cancer specific probes when compared to our chRCC data and TCGA methylation data for a small cohort of ccRCC and pRCC. This observation of frequent hypomethylation may be a distinguishing feature for the less malignant RCC subtypes. As the more malignant RCC subtypes, ccRCC and pRCC presented with hypermethylation within the majority of cancer specific probes (>80% cancer specific probes). This finding is consistent with the generally accepted observation that methylation is more

prevalent in malignant and aggressive tumours. Other recent publications on the methylation profile of RCCs include: a TCGA chRCC specific publication where chRCC 450K methylation was compared to that of ccRCC (Davis et al., 2014), a study using 27K Infinium methylation array of RCC subtypes (ccRCC, pRCC, chRCC and oncocytoma) (Ibragimova et al., 2013b), and a candidate gene methylation analysis between RCC sub groups (Costa et al., 2007; Dulaimi et al., 2004; Ellinger et al., 2011b). We identified pRCC (TCGA data) to present with the most cancer specific hypermethylated probes, it was calculated that pRCC presented with 3.4x more hypermethylated probes than ccRCC. This observation echoes the findings of McRonald et al., 2009, where Goldengate methylation analysis identified much higher numbers of methylated loci in a pRCC cohort compared to ccRCC.

Supervised clustering of hypermethylated probes, did differentiate the majority of oncocytoma and chRCC into groups with group cluster 1.1 and 1.2 predominantly oncocytoma while cluster 1.3 and cluster 2 were predominantly chRCC. This clustering analysis did identify one chromophobe sample that presented with widespread hypermethylation, a feature that may be reminiscent of a CIMP phenotype (Issa, 2004). However, further investigation and samples would be required to confirm such an observation.

This study has permitted the identification and characterisation of frequently (>30% of samples) hypermethylated and hypomethylated loci within a chRCC and renal oncocytoma cohort. When comparing the two histologies, we identified that chRCC presents with more frequent hypermethylation of CpG loci (416 probes) than the

oncocytoma subgroup (125 probes), which is consistent with chRCC being a malignant form of cancer compared to renal oncocytoma. Furthermore mapping of the hypermethylated and hypomethylated loci position in relation to gene, identified no obvious association between the methylation status and position in relation to genes, unlike a study conducted by Sandoval et al 2011, where hypomethylated probes were mainly identified in the gene body (38-40%), although Sandoval et al drew no conclusions from this observation, other studies have reported that loss of DNA methylation within the genes bodies may act as a mechanism combat gene over expression (Yang et al., 2014) When mapping the location of the hypermethylated and hypomethylated probes in relation to CpG islands distinct profiles were observed. Hypermethylated probes were more frequently located in GC rich regions of the CpG island, shores and shelves (>75% hypermethylated probes for both histologies) while hypomethylated probes predominantly resided in lone GC poor regions known as open seas (>65% hypomethylated probes for both histologies). This finding is consistent with the study by Sandoval et al., 2011, who identified hypermethylation to reside in 66-68% CpG islands, while identifying 67-70% of hypomethylated probe reside in open sea.

Differential analysis was conducted with the aim of identifying genes with abnormal methylation patterns that may provide a biomarker for differentiating between RCC subtypes. By identifying probes frequently hypermethylated in one histology compared to the other, we were able to identify 30 significantly differentially hypermethylated genes, 12 of which were located within promoter associated CpG islands, shores or shelves. A study by Irizarry et al., 2009 reported that differential methylation in colon

cancers was most prevalent in the CpG shores (up to 2kb away from the promoter) suggesting a that distal methylation may also regulate gene expression. For hypomethylated differential analysis, a stricter criterion could be applied due to the abundance of frequently hypomethylated probes identified (>1239 gene associated probes). This analysis identified 41 differentially hypomethylated genes, of which only 6 were located within promoter associated CpG shores and none within promoter CGIs. This is consistent with reports that hypomethylation is less frequent at regulatory regions of the genome (Ehrlich, 2009).

From our study, we identified two genes (*SPG20* and *NPHP4*) of interest to be significantly methylated in chRCC when compared to our oncocytoma cohort and TCGA ccRCC and pRCC data. ,A recent publication of the chRCC TCGA data, comparing multiple platforms, investigated the chRCC expression profile compared to normal and ccRCC. Investigation of the supplementary data identified a decrease in average fold difference (FD) of -2.04 *NPHP4* (FDR 1.73 e^{-14}) and -0.61 *SPG20* (FDR 1.59e^{-5}) in chRCC expression when compared to normal, while ccRCC expression was a FD of 1.74 *NPHP4* (FDR 2.27 e^{-51}) and -0.80 *SPG20* (FDR 3.61e^{-2}) more than chRCC. However, the CG probes identify in our investigation were not included in the lists significant probes within the TCGA supplementary data. (Davis et al., 2014).

SPG20 encodes a protein known as spartin, which has been shown to play a role in preventing cytokinesis arrest (Lind et al., 2011), while mutation within *SPAG20* are attributed to Troyer syndrome (an autosomal recessive spastic paraplegia syndrome characterised by spastic paraplegia, muscular atrophy and motor and cognitive delays)

173

(Tawamie et al., 2015). Furthermore methylation of *SPG20* has been suggested as a biomarker for early colorectal cancer, with methylation reported in 89% colorectal cancer and 78% adenoma (Lind et al., 2011), and has been shown to be highly methylated in Non-Hodgkin lymphoma (NHL), with frequent methylation identified in all NHL (75%- 100%) subgroups tested, with 0% in normal controls (Bethge et al., 2014).

NPHP4 has been identified as a cilia-associated protein nephrocystin-4, which is involved in the correct formation of primary cilium. *NPHP4* has been identified to have an antagonistic involvement in the canonical Wnt signalling through interactions with *inversin* (*NPHP2*) (Lienkamp et al., 2012). Within the pro-apoptotic Hippo pathway, NPHP4 has been identified to bind to LATS1 and inhibit the interaction and subsequent phosphorylation of the YAP/TAZ complex. Resulting in the transcription of growth factors such as *CTGF*. This suggests that NPHP4 is a negative regulator of the Hippo pathway, suggesting oncogenic tendencies (Habbig et al., 2011). Therefore, further investigation would be required to identify the role DNA methylation may have on *NPHP4* function.

Mutations in *NPHP4* have been linked with nephronophthisis, while a particular truncating mutation, c.2044C>T, p.R682*, has been identified to cause cerebello-oculo-renal syndrome with renal cysts and linked to male infertility (Alazami et al., 2014). Finally, other studies into methylation of *NPHP4* promoter have also identified high levels of methylation within chronic and rapid kidney disease (Wing et al., 2014).

The differentially methylated genes identified in this study may provide an insight into disease progression and could potentially be used as biomarkers to distinguish between

174

RCC subtypes. However, before such conclusions can be drawn, further functional analysis should be conducted.

Since the publication of our study, a further study by Ibragimova et al has been published with a similar aim of investigating the methylome profile of different RCC subtypes, including ccRCC, pRCC, chRCC and oncocytoma. Methylation analysis was conducted using the Infinium 27K array. This study identified 8 genes, to be differentially methylated in oncocytoma when compared to ccRCC, and 5 ccRCC genes to be differentially methylated in ccRCC. It was subsequently proposed that the methylation status of these genes could provide a gene panel, which would permit discrimination of renal oncocytomas from ccRCC. Regarding the comparison of oncocytoma and chRCC, methylation, only one gene *RAC2* was reported to be differentially methylated in oncocytoma samples. No observations were reported regarding the methylation of *SPG20* and *NPHP4* (Ibragimova et al., 2013b). Disparities between our data and this study are likely due to; differences within the tumour histopathology of the samples used, the different arrays (NPHP4 cg20383686 is only present on the 450K array), filtering criteria applied and different methylation thresholds (Ibragimova et al set differential methylation as a as β value <0.15 in all samples of an unmethylated histology but methylated (β value >0.45) in at least one sample for a separate histology). Highlighting the lack of consensus as to how to process this data, and what is biologically relevant.

Recently, the TCGA has published data on 66 chromophobe samples, investigating the genome wide methylation profile (450K analysis), expression profile (RNA-sequencing

and miRNA seq) and mutation and SNP profile (Whole genome sequencing and SNP array), to identify the genomic and the epigenomic landscape of chRCC (Davis et al., 2014). Within this study, Davis et al, also identified more global hypomethylation than hypermethylation when comparing chRCC to ccRCC. Furthermore, they identified distinct patterns in methylation corresponding to the lineage of the two cancer subsets, with chRCC derived from intercalated cells in the distal convoluted tubules (associated genes were unmethylated in chRCC and methylated in ccRCC) while ccRCC are derived from the proximal tubules (thus genes associated with this cell lineage were hypermethylated in chRCC and unmethylated in ccRCC).

Other studies investigating biomarkers, allowing ccRCC and oncocytoma subtypes to be distinguished, included an investigation in the SNP profile which observed that copy number alterations were more prevalent in chRCC than oncocytoma, and frequent deregulation mTOR signalling pathway in chRCC. Furthermore, expression analysis and immunohistochemistry distinguished three potential biomarkers: parafibromin (*CDC73*), aquaporin 6 (*AQP6*) expressed in oncocytoma and synaptogyrin 3 (*SYNGR3*) expressed in chRCC (Tan et al., 2010). A further study into differentially expressed biomarkers has identified HNF1b and S100A1 to be strong candidates for immunohistochemical differentiation between chRCC and oncocytoma, with significantly more frequent and intense staining within oncocytomas (HNF1b = 73% (P =0.001) and S100A1 = 80% P < 0.001 of oncocytoma samples presenting with moderate to strong staining). Moreover, no chRCC tested presented with positive staining for either protein (Conner et al., 2015).

Profiling the micro-RNA signature of each RCC histology has identified unique expression patterns that can accurately differentiate between normal, ccRCC, pRCC chRCC and oncocytoma. Algorithms have been created that differentiate subgroups based on expression of specific mirco-RNA, for example miR-200b and hsa-miR-139 had increased expression in chRCC when compared with oncocytoma, while hsa-miR-200c was increased in oncocytoma (Fridman et al., 2010; Youssef et al., 2011).

3.3.2 Limitation of this study

The major limitation of this study was the small cohort size. The cancers investigated in this study are rare and at the time of investigation only a limited number of samples were available, making it challenging to draw meaningful. Additionally, the size of the array limited the number of samples that could be analysed, as additional arrays would have been required to profile a larger cohort. This would not only effect the costing's of the experiment, but would also open up issues of variability between arrays, and thus further normalisation of the data would be required (Harper et al., 2013).

Another limitation is regarding the histology of the tumours samples. This study is reliant on the accurate differentiation and identification of the two RCC subtypes, chRCC and oncocytoma. As previously discussed (sections 1.4.2.1.3 and 1.4.2.1.4) distinguishing of the two subtypes can be challenging, relying on visual characterisation of the histopathology (Ng et al., 2014; Yusenko, 2010), which maybe further complicated by intratumour heterogeneity present in the majority of cancers (section 1.2.1) (McGranahan and Swanton, 2015; Tabassum and Polyak, 2015). Identification and

classification of the chRCC and oncocytoma histologies was conducted by our collaborates Dr G Kovacs and Dr M Yao, both of whom are experienced in the identification and classification of renal cancer histologies.

Furthermore, a limitation with the technology itself, is the chemical differences between the two Infinium probe technologies employed by the 450K. It has been identified that the two different technologies produced different β value ranges, with the Infinium II probe producing a narrower range of β values, skewing recordings at the extremes of the β value range (0 and 1). This observation suggests that the Infinium II technique is not as sensitive as Infinium I (Bibikova et al., 2011; Sandoval et al., 2011). However, Infinium argue that the differences between the two technologies should not affect reliability or reproducibility and that the array was designed for probes to be analysed individually across samples and not compared between the two technologies within the same sample (Illumina, 2012). Despite this statement, it is widely agreed that all data should be normalised prior to analysis, so that both techniques are representative of the same range of β values. For our data this was achieved by the use of the Lumi pipeline, (Du et al., 2008), however other packages such as BMIQ (Beta MIxture Quantile dilation) (Teschendorff et al., 2013) and Subset-quantile Within Array Normalisation (SWAN) (Maksimovic et al., 2012) are now also available (reviewed by Morris and Beck, 2015). Another limitation was raised by Harper et al, regarding the distribution of the CpG probes in relation to genes. It was identified that some genes may be represented by more CpG probes than others, generating a bias that results in some genes of interest being disproportionately represented (Harper et al., 2013).

Furthermore, the global nature of the 450K analysis results in huge data files that require bioinformatic interpretation. This analysis itself can limit the functionality of the genome wide methylation studies, with many probes excluded throughout the various stages of filtering. Another limitation was defining where to set the thresholds for methylation, and as previously mentioned, there was a lack of consensus as to what thresholds should be applied and should be regarded as biologically relevant (Ibragimova et al., 2013b; Sandoval et al., 2011). However, more recent studies are moving away from the threshold method to detect differential methylation instead are applying statistical methods employed by multiple pipelines to assist the filtering of the 450K data (Morris and Beck, 2015).

Regarding the to analysis of the 450K data, another limitation to be aware of is the misrepresentation of samples methylation profile due to the presence of imprinted genes, X chromosome inactivation and CpG probes associated with known SNPs. All of which may skew the global methylation profile and could be misrepresented as a true result (Chen et al., 2013). However, for our data, this limitation was overcome by omitting all associated probes.

3.3.3 Future work

Much of the methylation observed through high throughput studies such as this, is passenger methylation, defined as methylation that does not effect the expression of the gene, or dive the formation of the observed phenotype. This may be due methylation occurring at genomic locations not associated with a gene or gene regulation or due to

methylation be acquired at a gene that was previously repressed within the tumour sample (Kalari and Pfeifer, 2010; Roy et al., 2014). Therefore future work would include identifying if the differential methylation observed for *SPG20* and *NPHP4* was driver methylation or passenger methylation. Expression analysis by RT-PCR (section 2.7.2) should therefore be conducted on cDNA to identify if the presence of methylation inhibits the transcription of mRNA. Additionally it should be identified whether 5'Aza treatment to remove methylation (in cell lines) restores mRNA expression, to identify that it is the DNA methylation that is resulting is loss of mRNA expression and not other regulatory features within the cell. If the presence of methylation was confirmed to reduce mRNA expression, the next stage would be to knockdown the protein and investigate the effect that loss of the protein has on cell growth and identify if *SPG20* or *NPHP4* may play a role in tumourigenesis. Furthermore, other non chRCC subtypes should be tested to confirm that *SPG20* and *NPHP4* are unmethylated in separate independent cohorts.

Further investigations should also focused upon identifying if any of the differentially hypomethylation genes identified by comparing chRCC and oncocytoma hypomethylation, are also differentially hypomethylated when compared to TCGA pRCC and ccRCC. Further functional analysis should be undertaken on any identified genes, such as colony formation studies should be conducted. Transfecting cells to overexpress the gene of interest, to ascertain if expression promotes tumour growth. Although hypomethylation is fairly common in cancer, it is often overlooked. However, as DNA hypomethylation is predominantly associated with repeats elements, such as long

interspersed elements (LINE-1) repeats and tandem, features which are highly prevalent in the genome. It is suggested the hypomethylation permits activation of transposable elements by relaxing the intergenic chromatin structure, resulting in DNA rearrangements and promotion of tumourigenesis (Ehrlich, 2009; Lorincz et al., 2004). Moreover reports have been published suggesting global reduction in methylation, is a marker of pre-cancerous breast cancer and early stages of urothelial cancer (van Veldhoven et al., 2015; Wolff et al., 2010).

Finally, open sea and intragenic probes identified to be frequently hypomethylated and hypermethylated within chRCC and oncocytoma should be mapped to identify if the loci reside within enhancer regions (identified on Encode) or map to microRNAs, which have previously been reported to have a regulatory role lost in the progression of cancer (Ge et al., 2015; Lujambio et al., 2010; Su et al., 2015). This may identify other regulatory features which are deregulated by DNA methylation.

3.3.4 Final summary

This study aimed to identify a differential methylation profile that may be used to distinguish chRCC from oncocytoma. By utilising the high-throughput genome wide 450K array, the methylation status >96% of CpG sites in more than 99% of RefSeq genes were interrogated. Through this analysis it was identified that both chRCC and oncocytoma RCC subgroups, presented with low levels of global hypermethylation and high levels of global hypomethylation, epically when compared to the more malignant ccRCC and pRCC TCGA data. Furthermore two promoter CGIs mapped to *SPG20* and

NPHP4 were identified to be frequently methylated in chRCC with no methylation observed within the oncocytoma, ccRCC or pRCC cohorts. Further investigation would be required to confirm if the presence of methylation at these genes has any implication chRCC tumourigenesis and if these two genes may provide a suitable biomarker of chRCC. However, this line of investigation was unable to be continued in this thesis due to time constraints. This was because the data from the whole exome sequencing of oncocytoma samples had been returned and required analysis (chapter 5).

As previously mentioned this work was published in Epigenetics 8, 2013 by Slater, AA et al. Additionally this work was presented at a poster session at: Epigenetics Europe, Genomic Research Europe; 2012 September 4-5; Frankfurt, Germany by Slater, A. et.al.

Methylight analysis of RASSF gene family in sarcomatoid Renal Cell Carcinoma (sRCC)

4.1 <u>Introduction</u>

4.1.1 Methylight

Methylight is an adapted form of the Taqman® real-time PCR technique, specifically for use with bisulfite modified DNA. It is a highly sensitive and quantitative method developed to assess DNA methylation status within regions of interest.

Methylight works by combining primers specific for methylated bisulfite modified DNA with a highly specific Taqman® style probe also specific for methylated bisulfite modified DNA. The probe is designed with a fluorescence tag (FAM or VIC) at the 5' end of the probe and a quencher molecule (TAMAR or BHQ) at the 3' end of the probe. During PCR amplification the methylation specific probe is degraded by the taq polymerase, this releases the 5' fluorescence tag from being in close proximity to the 3' quencher molecule, allowing the fluorescent tag to emit a signal that is then detected as successful amplification of 1 strand of DNA. Methylation of the sample is quantified in relation to an *in-vitro* methylated control and a housekeeping gene (e.g. *Alu* or *GAPDH*) and displayed as Percentage of Methylated Reference (PMR) (section 2.3.2.2.1) (Buhmeida et al., 2011; Dallol et al., 2011; Selamat et al., 2011).

Unlike MSP where it is solely the primers that differentiate between the methylated and unmethylated sequence (Herman et al., 1996) or CoBRA and bisulfite sequencing, where the PCR product is interrogated for methylation (Xiong and Laird, 1997). Methylight identifies methylation within both the primers and the product (fluorescent probe), providing high specificity and accurately determining and quantifying the methylation

status (Coleman and Rivenbark, 2006; Eads et al., 2000; Trinh et al., 2001). Another key advantage to Methylight is the sensitivity of the assay, meaning that much less template DNA (as little as 10ng) is required for accurate analysis (Dallol et al., 2011; Ogino et al., 2006) and valuable patient samples can be preserved for future investigation. Additionally, Methylight can be run in a semi high-throughput manner permitting analysis of multiple samples at one time (Coleman and Rivenbark, 2006; Eads et al., 2000).

Methylight has since been used for many studies including quantifying the methylation status of *RASSF1A* in breast cancer (Buhmeida et al., 2011), and hepatocellular carcinoma (Xu et al., 2013) or identifying differential methylation between different stages of adenocarcinoma (Selamat et al., 2011). More recent studies have utilised Methylight to identify methylated genes associated with negative prognosis within rectal cancer (Exner et al., 2015). Furthermore Methylight has also proved invaluable for studying the methylation profiles for diseases such as myelodysplastic syndrome or aplastic anaemia where large DNA quantities of samples are rare and difficult to obtain (Li et al., 2015).

Further advances in Methylight technology are being focused on improving efficiency; Olkhov-Mitsel et al., 2014 reported a method to multiplex Methylight. By utilising probes specific to 4 different target genes with each probe possessing a different coloured fluorescent dye, selected to have sufficient spectral separation to avoid overlap. They successfully managed to simultaneously amplify, detect and analyse 4 genes (*IAPC*, *HOXD3 TGFBP2* and *ALU)* within one well, with high specificity and reproducibility

(Olkhov-Mitsel et al., 2014). Other approaches have been to adapt the novel droplet digital PCR (ddPCR) method for use within the Methylight system (Yu et al., 2015). The ddPCR is a method developed to allow high throughput PCR analysis. Samples are fractionated into multiple droplets containing template DNA and PCR reagents, PCR amplification then occurs in each individual droplet. By suspending the PCR reaction and template DNA in water-in-oil droplets, each droplet acts as a separate PCR reaction and amplification. This technique first described by Hindson et al., 2011 has been reported to allow standard Taqman analysis of ~2 million PCR reactions per 96 well plate. The adaptation of the technology to encompass Methylight primer and probe design has permitted the identification and quantification of alleles and infrequently methylated genes within normal colon tissue, which may serve as an early warning biomarker for colorectal cancer. Furthermore this technique is reported to identify infrequent methylation with 25-fold more sensitivity compared to the standard Methylight and with less variation between samples being recorded (Yu et al., 2015).

4.1.2 Sarcomatoid renal cell carcinoma (sRCC)

As described in section 1.4.2.1.5, sarcomatoid RCC is a characteristic feature of advanced RCCs rather than a district subset of RCC and is associated with poor prognosis and metastasis (Delahunt et al., 2013; Yan et al., 2015). The sarcomatoid features are usually diagnosed using core biopsies of the tumour, followed by visual analysis of the sarcomatoid-like cells. These areas are often densely populated with elongated, spindle-shaped cells, and usually exhibit micro-vascular invasion and necrosis (Shuch et al., 2012a) (Figure 4-1).
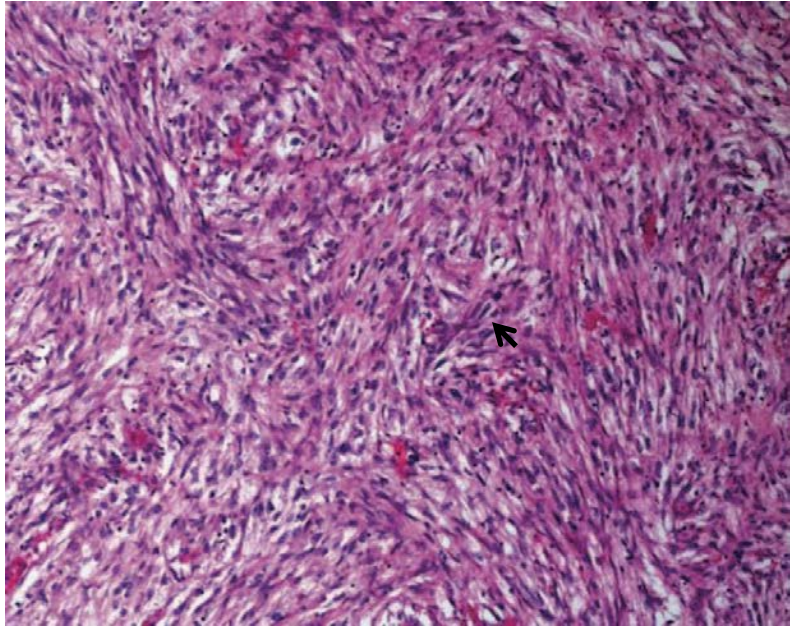
Figure 4-1: Histological image of sarcomatoid RCC.
Image shows the characteristic dense elongated spindle shaped cells in a sample derived from ccRCC with 70% sarcomatoid features (Black arrow highlights one example of sarcomatoid cells). Image from Shuch et al., 2012a.

It is predicted that sarcomatoid features present in 1-8% of all ccRCC tumours (Humphrey, 2012) and in a total of 16-20% of advanced renal cancers and can be derived form virtually any malignant RCC histology including pRCC, chRCC and collecting duct RCC (Shuch et al., 2012a; Yan et al., 2015)(Delahunt et al., 2013). However, the amount of sarcomatoid features can vary hugely within individuals from 1% to 100% of the tumour histology though this is not always reported (Kim et al., 2015).

4.1.3 RASSF Family

The RASSF family are a group of 10 tumour suppressor proteins (RASSF1-RASSF10) frequently identified to be downregulated in many cancers including RCC (Maher, 2013). The RASSF family of proteins have been associated with important regulatory roles in cell growth regulation via pro-apoptotic signalling, cell cycle arrest and the Hippo pathway (section 4.1.3.1) (Richter et al., 2009). Structurally, RASSF proteins are characterised by the presence of a RAS-association domain (RalGDS/AF-6) (RA domain), which facilitates the binding of RASSF protein to RAS or RAS-GTP (Agathanggelou et al., 2005). Although all RASSF family members contain the RA domain, the presence of other highly conserved domains (detailed below) allows for sub-classification of the family into the following two groups;

1) **Classical RASSF proteins**

   Classical RASSF proteins (RASSF 1-6) possess a Sav-RASSF-Hpo domain (SARAH domain), a domain that is highly associated with protein: protein interactions and protein heterodimerisation (Agathanggelou et al., 2005). The presence of a SARAH domain has been associated with the binding of RASSF proteins with the

Hippo pathway associated proteins SAV, MST1 and MST2. RASSF1A, RASSF2, RASSF3, RASSF4 and RASSF5, have all previously been identified to bind to the core hippo pathway proteins MST1 and MST. Promoting activation of the pathway, thus activating the downstream Hippo pathway signalling and driving apoptosis (Bao et al., 2011; Chan et al., 2013; Cooper et al., 2009; Fausti et al., 2012; Guo et al., 2007; Hwang et al., 2014; Iwasa et al., 2013; Matallanas et al., 2007; Romano et al., 2010; Shivakumar et al., 2002).Additionally RASSF3 and RASSF6 have been associated with p53 induced apoptosis by ubiquitinating and facilitating the degradation of p53 inhibitor MDM2 (Withanage et al., 2012) suggesting that RASSF proteins have more than one method of action for regulating cell growth.

2) N terminal RASSF proteins

The N terminal RASSF proteins (RASSF 7- 10) are characterised by their lack of a SARAH domain, and as yet their exact method of action and associated regulatory mechanisms still remains unclear (Underhill-Day et al., 2011a). However, reports have suggested that these N-terminal RASSF proteins may play a role in mitosis and cell cycle progression. For example RASSF7 has been reported to be required for correct microtubule and spindle formation during mitosis (Recino et al., 2010), while RASSF10 has been shown to be localised to the centrosomes during mitosis (V K Hill et al., 2011). Furthermore, it should be noted that the SARAH domain lacking RASSF7 has also been identified to bind the hippo pathway

protein MST1, and is predicted to incite activation of the hippo pathway as seen

with the classical RASSF proteins (Chan et al., 2013; Dittfeld et al., 2012).

Additionally, both classical and N-terminal RASSF proteins have been shown to lack

enzymatic activity. Thus, it is now widely regarded that they elicit Hippo pathway

activation by acting as scaffold proteins and bringing proteins such as MST1 and MST2

together to form the core complex of the Hippo pathway, (Donninger et al., 2007; Pfeifer

et al., 2010; Romano et al., 2014).

In cancer, the RASSF family of TSGs have been identified to be frequently

underexpressed, prompting investigation into the DNA methylation status. As previously

mentioned in section 1.5 frequent methylation of *RASSF1A* and *RASSF5* have been

observed multiple RCC subtyped including ccRCC, pRCC, chRCC and oncocytoma (Costa

et al., 2007; Dulaimi et al., 2004; Ellinger et al., 2011b; Maher, 2013; Morrissey et al.,

2001). *RASSF1* promoter methylation was first identified by Dammann et al 2000, in

small cell lung carcinomas (Dammann et al., 2000). Since this identification methylation

of *RASSF1A* has been reported in numerous cancers such as breast, glioma,

hepatocellular and adenocarcinoma (Hesson et al., 2007; Selamat et al., 2011), with

several studies linking hypermethylation to poor prognosis and advance cancer stage

(Buhmeida et al., 2011; Selamat et al., 2011).

Further studies investigating the promoter methylation of other members of the RASSF

family, have identified frequent methylation of *RASSF2, RASSF6* and *RASSF10* in multiple

cancers, including; pheochromocytoma (Richter et al., 2015), breast cancer (Hagrass et

al., 2013), colorectal cancer (Heyn and Esteller, 2012), thyroid cancer (Schagdarsurengin

189

et al., 2010), lung cancers (Grawenda and O'Neill, 2015) hepatocellular carcinoma (Dong et al., 2015; Xu et al., 2013), brain gliomas (Hesson et al., 2004) and leukaemia (Hesson et al., 2009) to mention but a few. With other studies linking the downregulation of the *RASSF* TSG family with increased and unregulated transcription of oncogenic genes such as *AREG,* which can lead to tumourigenesis (Ahn et al., 2013),

4.1.3.1 *Hippo pathway*

The Hippo pathway is frequently deregulated in cancer. Originally characterised in Drosophila, the Hippo pathway is a highly conserved pathway with important roles in the regulation of organ size, cell proliferation, cell cycle and cell death (Bao et al., 2011; Pan, 2010)(Yabuta and Nojima, 2013). With activation of the Hippo pathway, leading to apoptosis and cell death (Hong et al., 2014).

In brief, the Hippo pathway is formed of a core cascade of proteins, Starting with a complex of serine/threonine kinases (MST1, MST2 and SAV1), this complex phosphorylates a second complex (formed of LATS1, LATS2 and a scaffold protein MOB1A) which in turn phosphorylates the oncogenic transcriptional co-activator complex YAP/TAZ. Preventing its transition into the nucleus and thus preventing gene transcription of pro-growth genes such as cyclin E, AREG and CTGF (Johnson and Halder, 2014; Meserve and Duronio, 2015; Pan, 2010) (Figure 4 2).
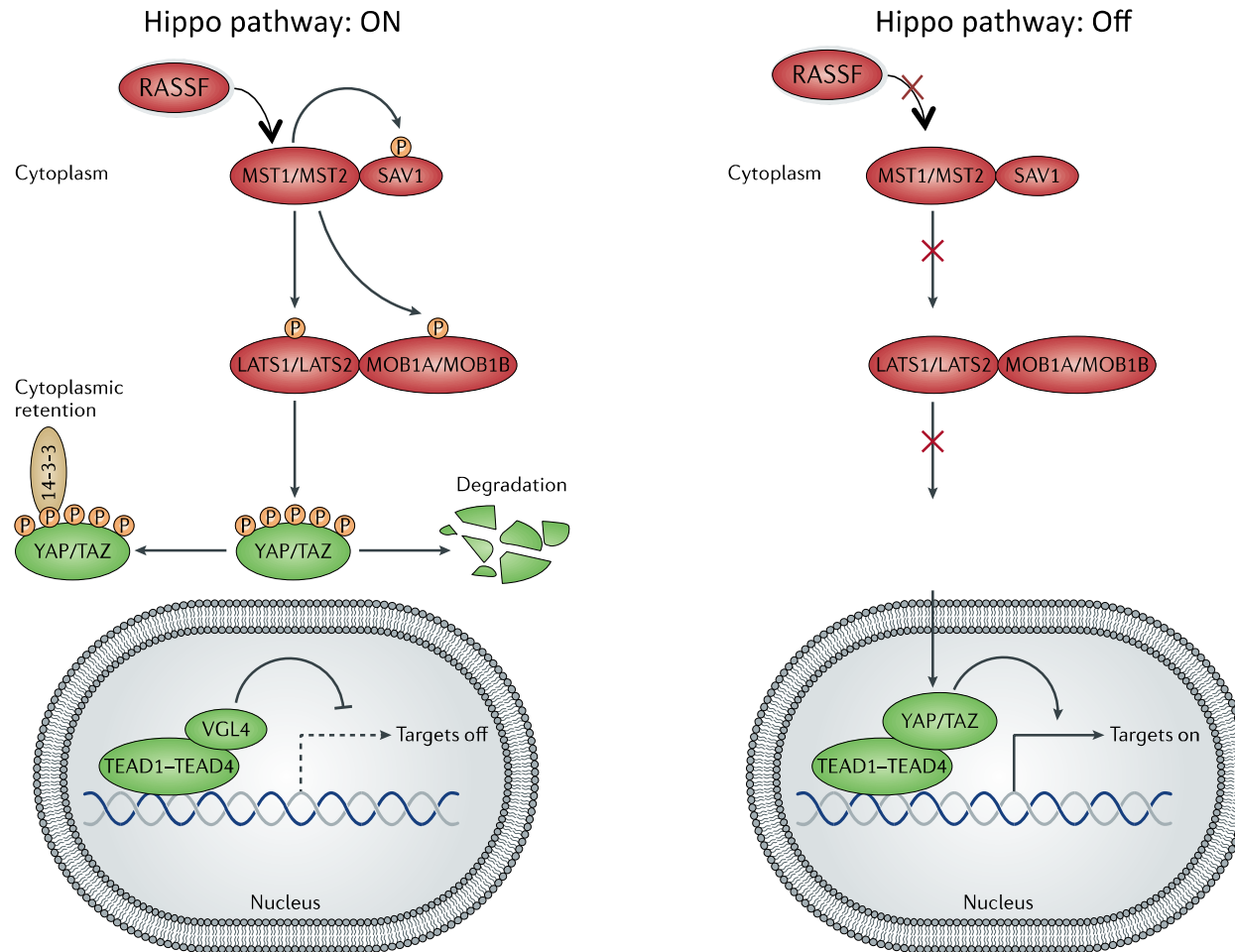
Figure 4-2: Schematic of the Hippo pathway.
Schematic of the mammalian hippo pathway showing the basic core pathway when activated (ON) and un-activated (OFF) and Green represents proteins that promotes activity of YAP/TAZ and red, identifies proteins which inhibit YAP/TAZ activity. Image is adapted form Johnson and Halder, 2014.

Additionally, phosphorylated YAP/TAZ, has been associated with pro-apoptotic functions of p73, while LATS2 causes reduced degradation of P53 through interactions with MDM2 (a known P53 suppressor) (Harvey et al., 2013). These observations highlight the complexity of the Hippo pathway and some of the mechanisms of action. As previously mentioned (section 4.1.3) the RASSF family of proteins have been identified to play an important role as scaffold proteins promoting the interaction of MST1 and MST2, and subsequent activation of the downstream components of the Hippo pathway signalling (Bao et al., 2011; Matallanas et al., 2007; Romano et al., 2010). The role of RASSF1A in the Hippo pathway is the most frequently studied, it has been identified that RASSF1 can dissociate MST1 and MST2 from inhibitory proteins such as Raf-1, and by forming homo and heterodimers (with other RASSF family members i.e. RASSF5) promotes the formation of the MST1/ MST2 complex and Hippo pathway activation (Bao et al., 2011; Matallanas et al., 2007; Romano et al., 2010),

The role of the Hippo pathway in cancer is a common focus of studies, especially since transcription of growth factors such as *CTGF* have been associated with chemotherapy resistance (Johnson and Halder, 2014). Mutational analysis of the Hippo pathway has identified that somatic mutations within the core Hippo cascade proteins are rare (Harvey et al., 2013; Morris et al., 2011). However, upstream activators and regulators of the pathway such as the tumour suppressor *NF2* have been observed to possess a high frequency of somatic mutations (>5% samples) in many cancers including glioblastoma, pituitary cancers and Schwannoma haemangioblastoma. Such mutations can result in

abnormal signalling of the Hippo pathway (Cancer.sanger.ac.uk, 2015; Harvey et al., 2013; TCGA, 2008).

Though mutations of the core Hippo pathway proteins is uncommon, frequent DNA methylation resulting in reduced expression of MST1/ MST2 and LATS1 has been reported in ccRCC (Chen et al., 2014), and soft tissue sarcomas, neurogenic sarcoma and liposarcomas (Seidel et al., 2007).

Furthermore, the frequent loss of expression of the *RASSF* family of TSG through DNA methylation reported in many cancers including RCC (section 4.1.3) has been shown to correlate with decreased activation of the Hippo pathway leading to a decrease in apoptosis (Ahn et al., 2013).Moreover, a recent study by Schütte *et al* 2014, has also identified that aberrant signalling of the Hippo pathway is associated with more aggressive, invasive and metastatic ccRCC (Schütte et al., 2014). Providing further evidence that the Hippo pathway is one of the key pathways defective in RCC.

4.1.4 Study's aim

As previously described (section 1.4.2.1.5 and section 4.1.2), sarcomatoid RCC is a histological feature observed in between 1-8% of all ccRCC tumours and associated with advanced, aggressive and metastatic RCC and poor prognosis (Delahunt et al., 2013; Humphrey, 2012; Mian et al., 2002; Yan et al., 2015). However, little research has been conducted into epigenetic profiles of sRCC, with a recent PubMed search identifying that most sRCC based papers, are reports on histological observations and case studies [NCBI PubMed search, Keywords; Sarcomatoid Renal Cell Carcinoma sRCC; conducted on 14th

August 2015]. We therefore propose to undertake a candidate gene methylation study to identify methylation profiles of sRCC.

Due, to the identification of frequent methylation of the *RASSF* family of TSG in non sarcomatoid RCCs (Costa et al., 2007; Dulaimi et al., 2004; Maher, 2013; Morris et al., 2011). The observation that high levels of *RASSF1A* hypermethylation, were associated with advanced cancer grade and patient prognosis (Buhmeida et al., 2011; Selamat et al., 2011). In addition to the identification of RASSF family proteins playing a key role in the activation of the pro-apoptotic Hippo pathway (Bao et al., 2011). A pathway that has been identified to be deregulated in several cancers, including advance and aggressive ccRCC (Johnson and Halder, 2014; Schütte et al., 2014). It was determined that the *RASSF* family of genes were good candidates, in which to investigate DNA methylation within a cohort of aggressive and metastatic sRCC.

We therefore propose to undertake a candidate gene approach using Methylight analysis to characterise the methylation profile of the frequently methylated *RASSF* family of genes members *RASSF1a, RASSF2, RASSF6* and *RASSF10* in a cohort of sRCC samples derived from ccRCC. With further investigation conducted to ascertain if there is any association between methylation status and clinical data (objective 2 section 1.7).

4.1.5 Results

4.1.6 Methylation profile of RASSFs via Methylight

The methylation status (section 0) of *RASSF1a, RASSF2, RASSF6 and RASSF10* was determined in 31 Japanese samples of sRCC derived from ccRCC previously isolated and classified by Dr M Yao (section 2.1.2) and compared to normal non-diseased kidney

(Figure 4-3). However, data regarding the percentage of sRCC features within the tumours was not supplied. As previously described, *RASSF1a, RASSF2, RASSF6* and *RASSF10* are frequently methylated in several cancers; therefore, we undertook a targeted approach to investigate the methylation status within the rare and aggressive sRCC.

Methylight is a highly sensitive technique, therefore in concordance with the literature a PMR of >10 was considered to represent the presence of methylation in a sample (Buhmeida et al., 2011; Dallol et al., 2011). Of the candidate genes investigated, we identified frequent methylation of *RASSF2* and *RASSF10* promoter regions in sRCC samples when compared to normal kidney. *RASSF2* was identified to be methylated in 25.8% (8/31) of sRCC samples, however this was not statistically significant when compared to *RASSF2* methylation in normal kidney DNA (t-test for unequal variances P>0.5)

RASSF10 on the other hand was identified to be significantly methylated when compared to normal kidney DNA (t-test for unequal variances P<0.005). With a *RASSF10* methylation identified in a total of 48.4% (15/31) of sRCC samples. No methylation was observed within the normal kidney samples for either of these two candidate genes.

Analysis of the range of PMR values identified large variation for both *RASSF2* and *RASSF10*. For *RASSF2* the PMR values observed, ranged from 0% to 100.5% (with a median PMR of 0% and a mean PMR of 14.9%). For *RASSF10* the PMR values observed, ranged from a 0% to 116.4% PMR (with a median PMR of 9.7% and a mean PMR of 17.0%).
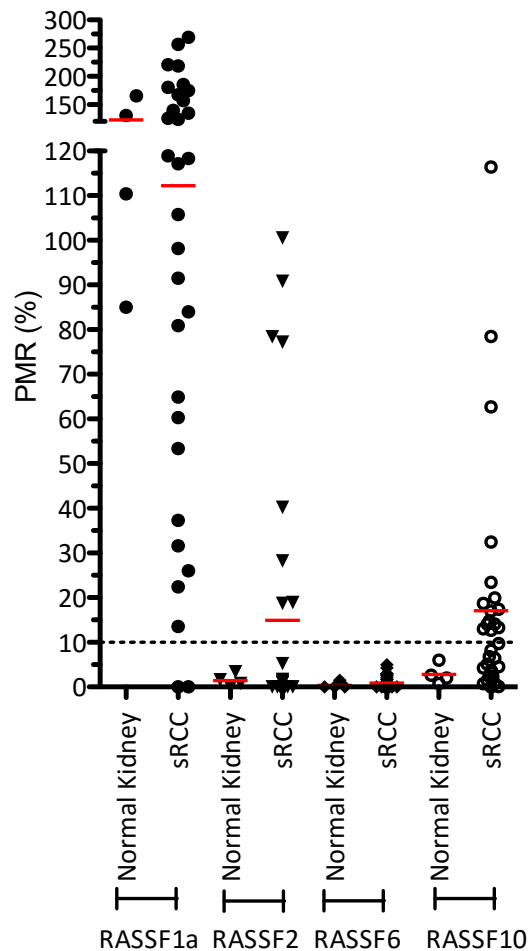
Figure 4-3: Methylight analysis of *RASSF1a, RASSF2, RASSF6* and *RASSF10*
Results of Methylight analysis investigating *RASSF1a, RASSF2, RASSF6* and *RASSF10* promoter regions, conducted on sRCC and compared to normal non-diseased kidney. Methylation is displayed on y-axis as an average PMR (n=3) for each sample, overall mean PMR is depicted by the red line. Samples with PMR >10 were deemed methylated, depicted by dashed threshold line.

RASSF1a presented with high levels of methylation in both the sRCC samples (median PMR = 117.1%, mean PMR = 112.2%, with a PMR range of 0% to 296% PMR) and the normal kidney (median PMR= 120.5% mean PMR = 122.8% and a PMR range of 85% to 165.2%), therefore the methylation was not cancer specific.

RASSF6 presented with no PMR value >10 in sRCC or normal kidney (sRCC samples presented with: median PMR=0%, mean PMR = 0.9%, and a PMR range of 0% to 5%) and was therefore defined to be unmethylated in this cohort of sRCC .

It was occasionally observed; particularly for *RASSF10* and *RASSF1a*, that a PMR value exceeded 100%. Selamat et al., 2011 suggesting this may be due to the *in-vitro* methylated control not being fully methylated at the particular region investigated, but is still representative of high methylation within the region investigated.

4.1.7 Association with Clinical Phenotype and Survival

To investigate if there is any association between the clinical data available for the sRCC (Table 2-1). Investigation into patient survival and the methylation status of *RASSF2* and *RASSF10*, Kaplan-Meier survival analysis was conducted using Mantel-Cox analysis (Figure 4-4). No significant difference was identified between survival rate and methylation status (methylated samples had a PMR>10%) for *RASSF2* (P>0.8) or *RASSF10* (P>0.5). Further investigation examined a possible association between methylation status of *RASSF2* and *RASSF10* in sRCC and the occurrence of metastasis (defined according to the M stage of the TNM score (section1.4.1) (Figure 4-5). Here we identified a potential trend, between and the methylation status of *RASSF10* (P=0.04 Pearson Chi-Squared test) and occurrence of metastasis.

A. Survival of RASSF2

B. Survival of RASSF10
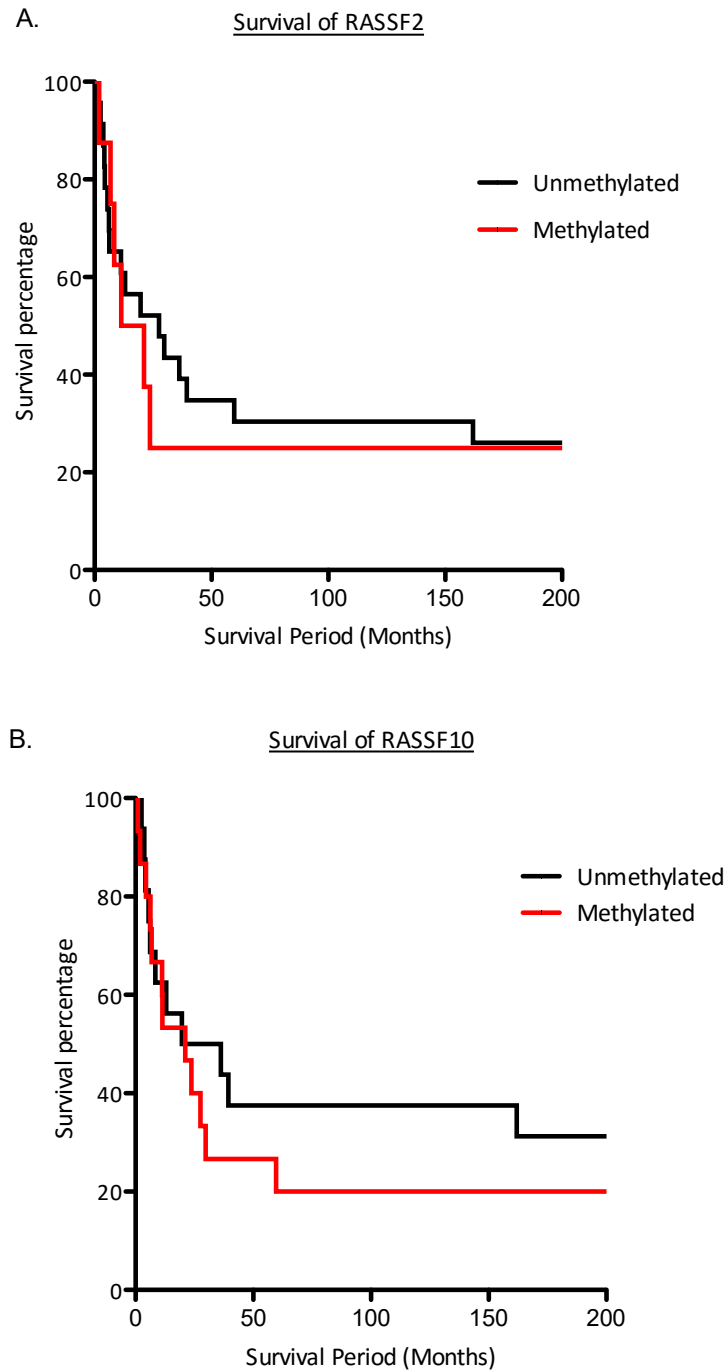
Figure 4-4: Kaplan-Meier survival curves

Kaplan-Meier survival curves of A) *RASSF2* and B) *RASSF10*, comparing patients' survival (months) to methylation state (Methylated (Red); Unmethylated (Black). No significant difference in survival between patients with methylation and those that were unmethylated was observed (Mantel-cox analysis; RASSF2 P=0.81; RASSF10 P=0.50)

A. RASSF2

B. RASSF10

■ Unmethylated samples   ▦ Methylated samples
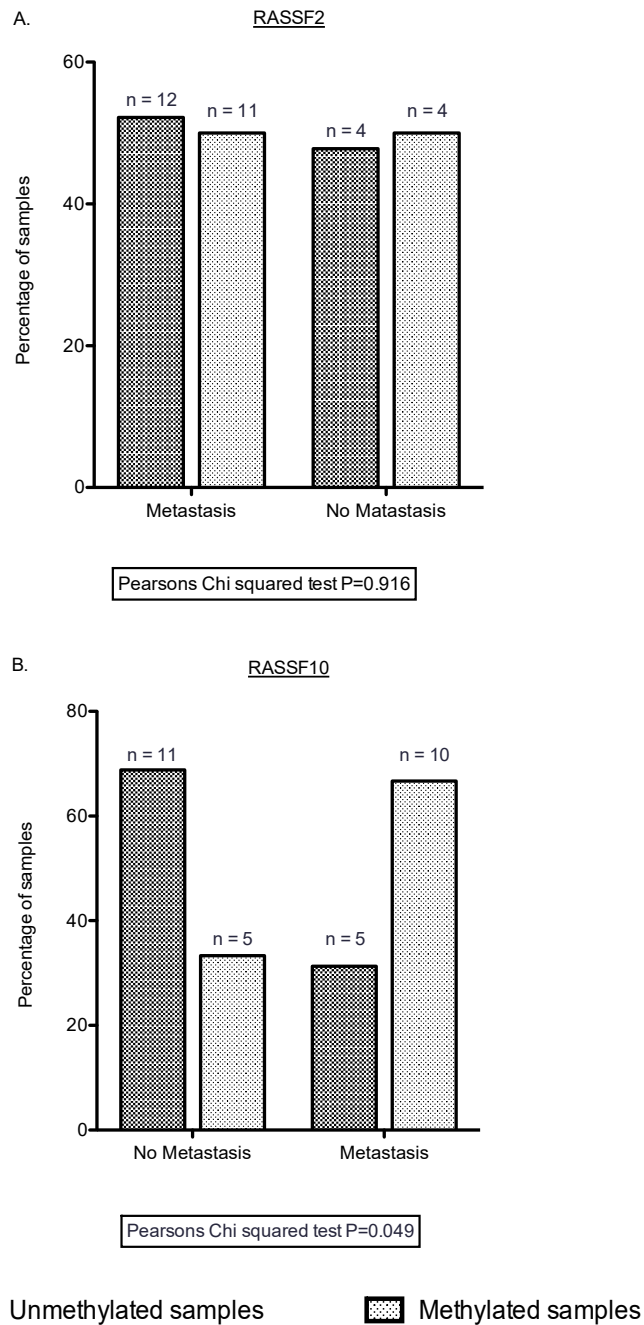
Figure 4-5: Association between methylation status and metastasis

The association between methylation status and metastasis is displayed. Bars represent the percentage of sample in which metastasis was observed in methylated and unmethylated samples., no significant association P=0.916 for **A)** *RASSF2.* **B)** For *RASSF10* metastasis was observed significantly more often (P=0.049) in methylated samples compared to unmethylated samples. Statistics were conducted using Pearson's Chi-Squared.

It was observed that there was a higher occurrence of metastasis in samples presenting with methylation of *RASSF10* (Figure 4-5b). This trend was not observed for *RASSF2* methylation (Figure 4-5a). No further clinical data was available as to the location of the metastasis. No further trends or associations were identified when comparing between methylation state, size of tumour, patient age or gender, nor were any further observations made when comparing clinical data with methylation status of *RASSF2* and *RASSF10* combined (Table 2-1).

4.1.8 RASSF10 Methylation in ccRCC

As previously mentioned the sRCC samples investigated within this study are derived from ccRCC, and sRCC is identified as a late stage and aggressive tumour. We therefore hypothesised that frequent *RASSF10* methylation maybe a potential feature of advanced late stage aggressive tumours.

To test this hypothesis, firstly, we investigated the methylation of RASSF10 in all paired normal TCGA ccRCC samples from the Illumina 450K data [correct as of 1st March 2013] (clinical data shown in Table 2-11). Identified no overall methylation of *RASSF10* associated probes, however it should be noted that the CpG probes of the 450K array does not map to the same region of the *RASSF10* CpG island which is interrogated by the Methylight primers, so further validation of the methylation status of *RASSF10* in ccRCC is required (Figure 4-6).

Probe	1 cg08148891	2 cg19665644	3 cg20918243	4 cg09596429	5 cg05817758	6 cg08876434	7 cg25344734	8 cg26005578	9 cg05095158	10 cg22740547	11 cg07600636	12 cg00326127	13 cg20098471
Average β Value — N	0.12	0.01	0.07	0.10	0.07	0.04	0.06	0.02	0.02	0.01	0.09	0.13	0.13
Average β Value — T	0.14	0.02	0.08	0.10	0.10	0.05	0.07	0.02	0.02	0.02	0.13	0.19	0.31
Frequency of methylation (β>0.5) — N	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)
Frequency of methylation (β>0.5) — T	0.6% (1/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	0.0% (0/160)	1.3% (2/160)	1.3% (2/160)	18.1% (29/160)

(160 N and 160 T samples)

Figure 4-6: TCGA 450K analysis of *RASSF10*

Methylation of *RASSF10* from TCGA 450K methylation data for 160 ccRCC with matched normal. Methylation status of each probes is shown as overall average of β values (top) and summary of the number of samples that were methylated as a percentage and frequency (bottom). Methylation was defined as β value >0.5.

450K probes (red) are mapped to a schematic of the *RASSF10* gene (blue) and CPG island (Green). Black line depicts the region covered by bisulfite sequencing and purple arrow depicts the region investigated by Methylight.

To further identify whether *RASSF10* methylation was a feature of advanced stages of RCC, a further Methylight study investigating the methylation status of *RASSF10* was conducted on an additional cohort of 20 Japanese ccRCC of various stages (6 = UICC stage I, 3 = UICC stage II, 6 = UICC stage III and 5 = UICC stage IV) (samples are separate from those analysed in chapter 3) (Figure 4-7). We identified 30% (6/20) of ccRCC samples presented with *RASSF10* methylation (PMR>10). The methylation was distributed across the 4 stages of ccRCC as follows;

 UICC stage: I 16.6% of samples (1/6 samples methylated) (average PMR=6)

UICC stage II :0% of samples (0/3 samples methylated) (average PMR=0)

UICC stage III: 16.6% of samples (1/6 samples methylated) (average PMR=3.7)

UICC stage IV: 66.6% of samples (4/6 samples methylated) (average PMR=27)

Moreover, statistical analysis suggested UICC stage IV to be significantly more methylated than stage I (t-test P=0.043), but not significantly different from the sRCC cohort (t-test P=0.28), when analysed using a student t-test for unequal variance. However, analysis using Fishers exact test identified no significant difference between any stages or ccRCC and sRCC (P>0.05). From this Methylight study we propose that there maybe a trend for increased frequency of *RASSF10* methylation in later ccRCC stages, when visually analysing the data (Figure 4-7), however the small cohort size means we lack the power to statistically confirm this observation.

Figure 4-7: Methylight analysis of *RASSF10* in ccRCC
The methylation status of ccRCC samples separated according to stage is displayed as PMR and compared to normal non-diseased kidney. Black line depicts the overall mean PMR for each stage.
Samples with PMR >10 were deemed methylated, as depicted by dashed threshold line. Student t-test * P<0.05

4.1.9 Confirmation of Methylight

Validation of the results obtained via the sensitive Methylight technique was conducted via single colony sequencing (section 2.3.4) on a small selection of sRCC (n=3) and ccRCC (n=3) samples to ascertain the methylation status of each individual CpG dinucleotide. Primers were selected to sequence a region of the *RASSF10* CpG island that also encompassed the CpG dinucleotides interrogated via the *RASSF10* Methylight primers.

As displayed in Figure 4-8 the region of the *RASSF10* CpG island sequenced is shown as a schematic against the gene. For the sample sequenced, the methylation status of each CpG dinucleotide is also displayed either as black (methylated) or white (unmethylated), with each row representing an individual allele.

For each sample, the methylation index (MI) for the entire region was calculated. Additionally, the MI for the CpG dinucleotides interrogated by the Methylight assay was also calculated and compared to the Methylight PMR for each sample.

The MIs calculated for the Methylight region were identified to be similar to the PMR values recorded for each sample in both histologies, thus validating the Methylight technique. Additionally, it should be noted that the Methylight region MI and PMR were also similar to the overall MI for the whole region sequence, suggesting that the 'snapshot' of methylation status provided by Methylight, is a valid representation of the methylation status of the wider CpG island.
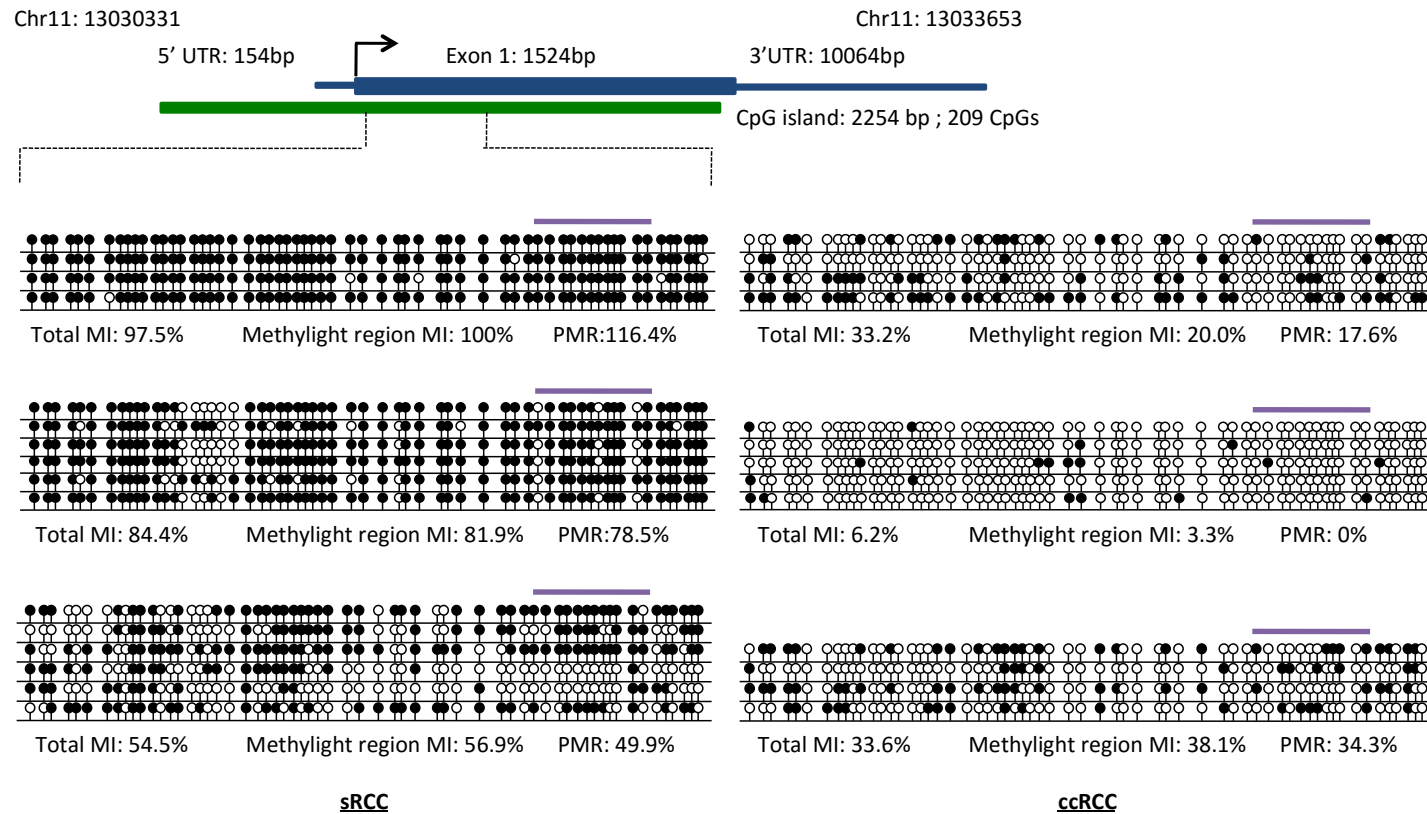
Figure 4-8: Validation of the Methylight PMR

The methylation status of *RASSF10* NM_001080521 was confirmed by single colony sequencing in 3 sRCC sample and 3 ccRCC. Each block represents a different sample with each row an individual allele, each circle depicts a CpG loci location and colour depicts methylation status; methylated (black) or unmethylated (white). The region investigated within the Methylight study is shown by the purple line above each sample. The total methylation index (MI) for the whole region single colony sequenced together with the recorded PMR and the MI for the Methylight region is shown. Schematic of the *RASSF10* gene (blue), CpG island (green) and the region investigated by Single colony sequencing (dashed lines) are shown.

4.1.10 Mutation Screen of RASSF10

Further investigations were conducted in order to establish if any other genetic features of *RASSF10* were present in the 31sRCC samples. The first approach was to establish if *RASSF10* was frequently mutated in sRCC and so the whole coding region of the single exon gene was examined via Sanger sequencing (section 2.5). No mutations were identified, within the 31 samples except for a known SNP, rs55909156 G>C at location 11:13010964 resulting in a non-synonymous change of p.C463S (Minor allele frequency (MAF) of 0.08 (8%) assessed via the 1000 genome project).

4.1.11 Loss of Heterozygosity of RASSF10

Loss of heterozygosity (LOH) is a common feature in cancer, particularly within the widely accepted two-hit paradigm (section1.2.2.2). Having found no somatic mutations, our attention was focused on identifying if any LOH may be present for RASSF10 in sRCC. LOH was investigated using a known size standard and primers specific for microsatellites flanking the RAFFS10 gene that were previously characterised to have a discernible range of product sizes (± 25bp) and a high probability of being heterozygous (> 0.6 probability) (section 2.2.7). LOH was detected by identifying if the primers both upstream and downstream of the GOI were homozygous, while the matched normal remained heterozygous. However, in 31 sRCC and matched N samples tested no LOH was observed.

4.1.12 Expression analysis

Expression analysis was then undertaken on renal cancer cell lines (ccRCC derived) to confirm that methylation of *RASSF10* corresponds to reduced expression of mRNA. Firstly, the methylation status of *RASSF10* in the ccRCC cell lines was established via CoBRA (Figure 4-9a). It was identified that for the RCC cell lines RCC11, UMRC3, A498 and A704 cell lines *RASSF10* was methylated while the KTCL140 and SKRC45 cell lines were unmethylated.

In order to determine the effect of DNA methylation on expression, the aforementioned cells lines were cultured by Mr D. Gentle in the presence of de-methylating agent 5-aza-2'-deoxycitidine (5-aza) (section 2.6.4), with a second untreated culture used as control. Following 4 days in culture RNA was extracted (conducted by Mr Dean Gentle)

Due to the single exon nature of *RASSF10* all RNA was subjected to an additional DNase clean up step, to remove residual DNA, prior to RT-PCR generation of the cDNA library and examination of *RASSF10* expression (section 2.7.2).

For the methylated cell lines, removal of methylation via 5-Aza treatment resulted in increased *RASSF10* mRNA expression when compared to the untreated controls. For the unmethylated cell lines (SKRC45 and KTCL140), demethylating treatment had no visible effect on the expression of *RASSF10* (Figure 4-9b). We therefore concluded that *RASSF10* methylation inhibits expression of the mRNA, however removal of the DNA methylation restored expression of the *RASSF10*.
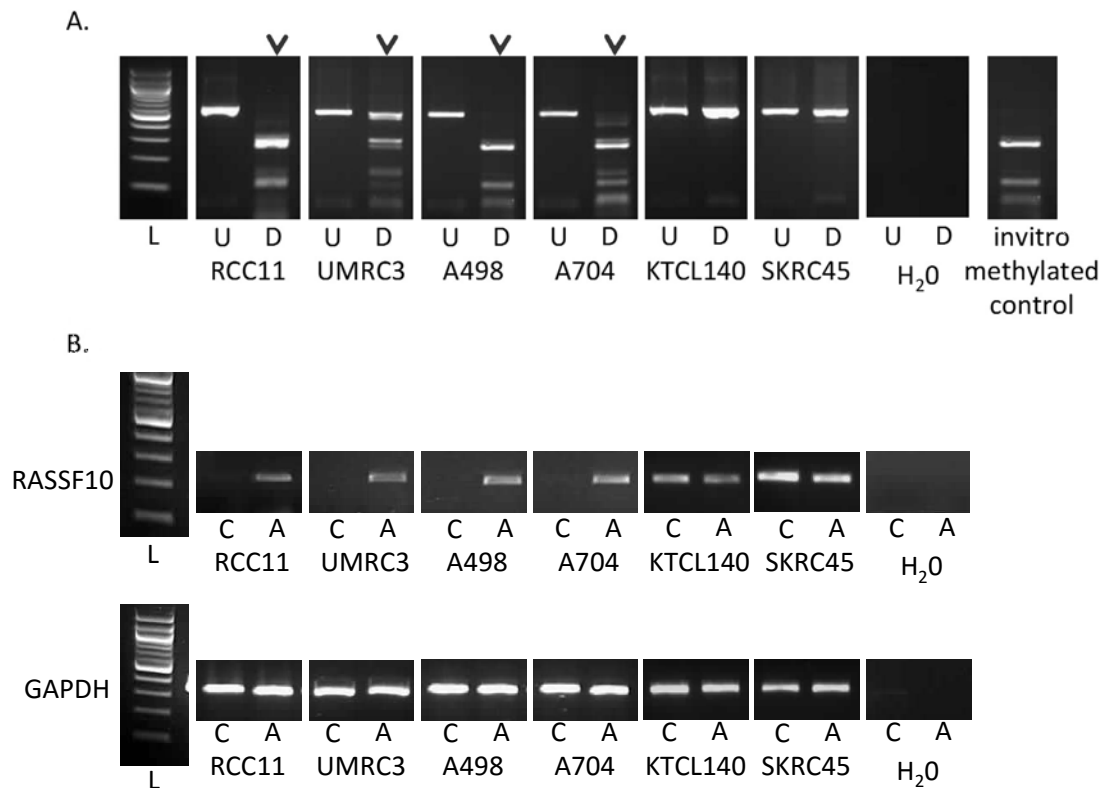
Figure 4-9: RASSF10 expression analysis
Methylation status and expression of *RASSF10* in kidney cancer cell lines. **A)** CoBRA analysis identifies *RASSF10* methylation in kidney cancer cell lines; samples are displayed as Undigested (U), compared to BSTU1 digested (D) arrows depict samples that are methylated. **B)** RT-PCR examining *RASSF10* mRNA expression between control (C) and 5-Aza treated (A) kidney cancer cell lines. Methylated cell lines did not express *RASSF10* mRNA. Expression of *RASSF10* mRNA was restored upon removal of DNA methylation after 5-Aza treatment. Unmethylated cell lines showed no difference in expression between control and 5-Aza treatment. *GAPDH* was included as a control for expression.

4.1.13 Soft agar assay of RASSF10 knockdown

At the time of this study little was known as to the function and method of action of RASSF10 (Underhill-Day et al., 2011b). Due to the high homology with other RAS binding proteins of the RASSF family, it had previously been predicted that *RASSF10* is a TSG. Studies in other cancers have previously identified frequent methylation of the *RASSF10* promoter associated in reduced expression of the *RASSF10* mRNA. Further *in-vitro* studies identified an increase in cell growth when *RASSF10* was knocked down in cell lines, a characteristic feature of TSGs (Hesson et al., 2009; V K Hill et al., 2011) .

To test that loss of RASSF10 results in increased cell growth within kidney cancers, we performed an anchorage independent growth assay (soft agar assay) in ccRCC cell lines (section 2.6.8). The use of soft agar assays is considered to be a more biologically relevant technique for investigating potential carcinogenic features such as knockdown of TSGs, as it allows for anchorage independent growth of the cell, a hallmark of carcinogenesis (Borowicz et al., 2014). *RASSF10* was knocked down (KD) via siRNA transfection in cell lines previously identified to express *RASSF10* (SKRC45 and KTCL140) (section4.1.12). Soft agar analysis of RASSF10 knockdown via siRNA (RASSF10-KD) in SKRC45 and KTCL140 cell lines resulted in a significant increase in colonies formation for both cell lines. For SKRC45 (Figure 4-10), negative control nonsense-siRNA (-ve Control) produced an average 102 colonies (n=5), while RASSF10-KD produced an average of 250 colonies (n=4, P = $2.39e^{-05}$). KTCL140 -ve control produced an average 134 colonies (n=5) while RASSF10-KD produced an average of 201 colonies (n=6 P=0.0026).

Western blots were conducted to analyse and quantify protein expression to confirm that siRNA KD of RASSF10 had resulted in reduced protein levels (Figure 4-11).

Normal convention would be to validate and control for off target effects of siRNA knockdown by performing a second knockdown of *RASSF10* and soft agar assay, repeating the experiment using a different siRNA targeted for *RASSF10.* However, in this case the siRNA used had previously been validated in our lab and compared to a second siRNA in U343 and H4 cell lines (V K Hill et al., 2011).
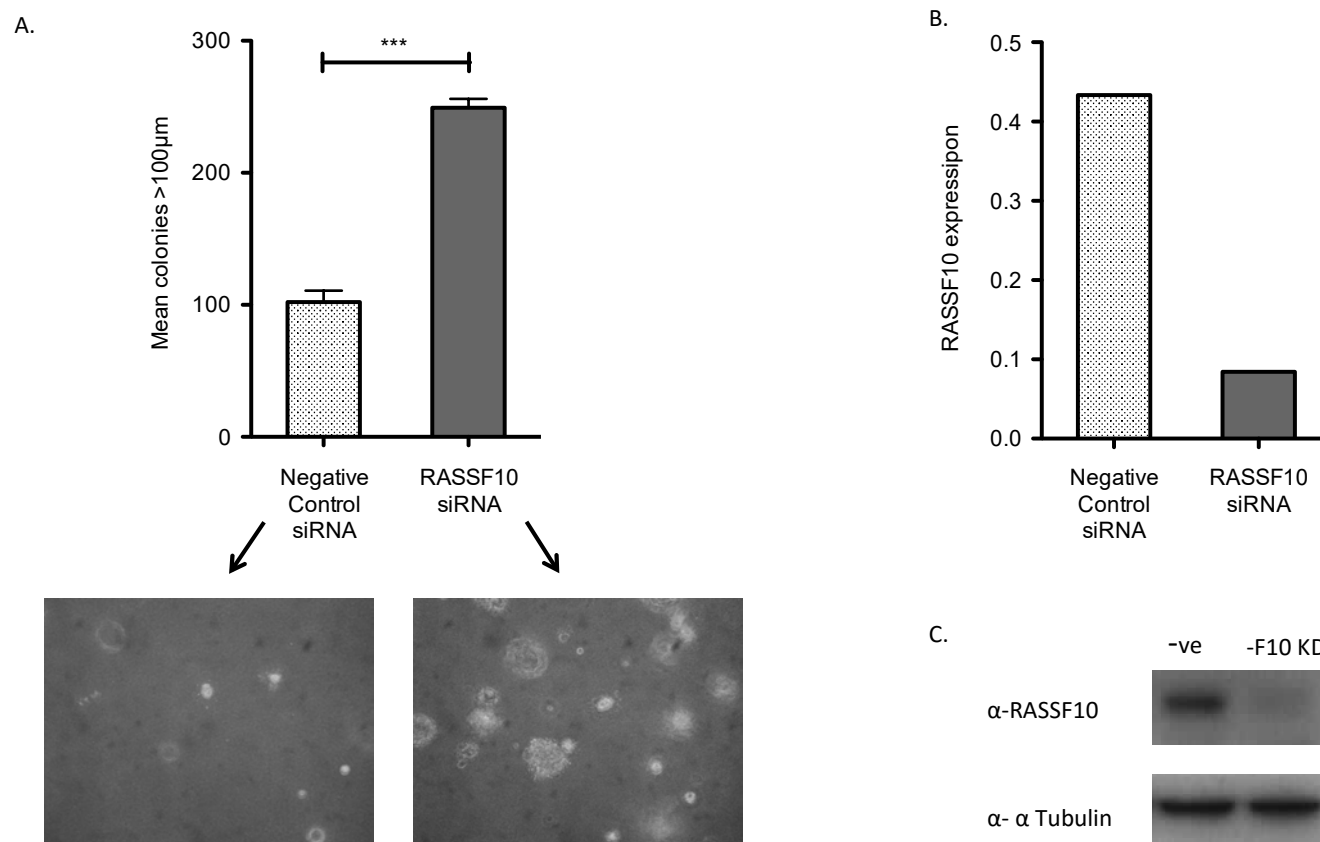
Figure 4-10: RASSF10 effect on cell growth in SKRC45

RASSF10 was knocked down in SKRC45 kidney cancer cell line by siRNA. **A)** Soft agar growth assay shows that knockdown of RASSF10 significantly increased colony formation and size. (t-test P=< 0.0001). Bar chart displays average colony count across repeats (n=3) + SEM. **B)** Quantification of western blot and **C)** Western blot image confirmed siRNA knockdown of RASSF10 correlates with knockdown of the protein. α-tubulin was included as loading control.
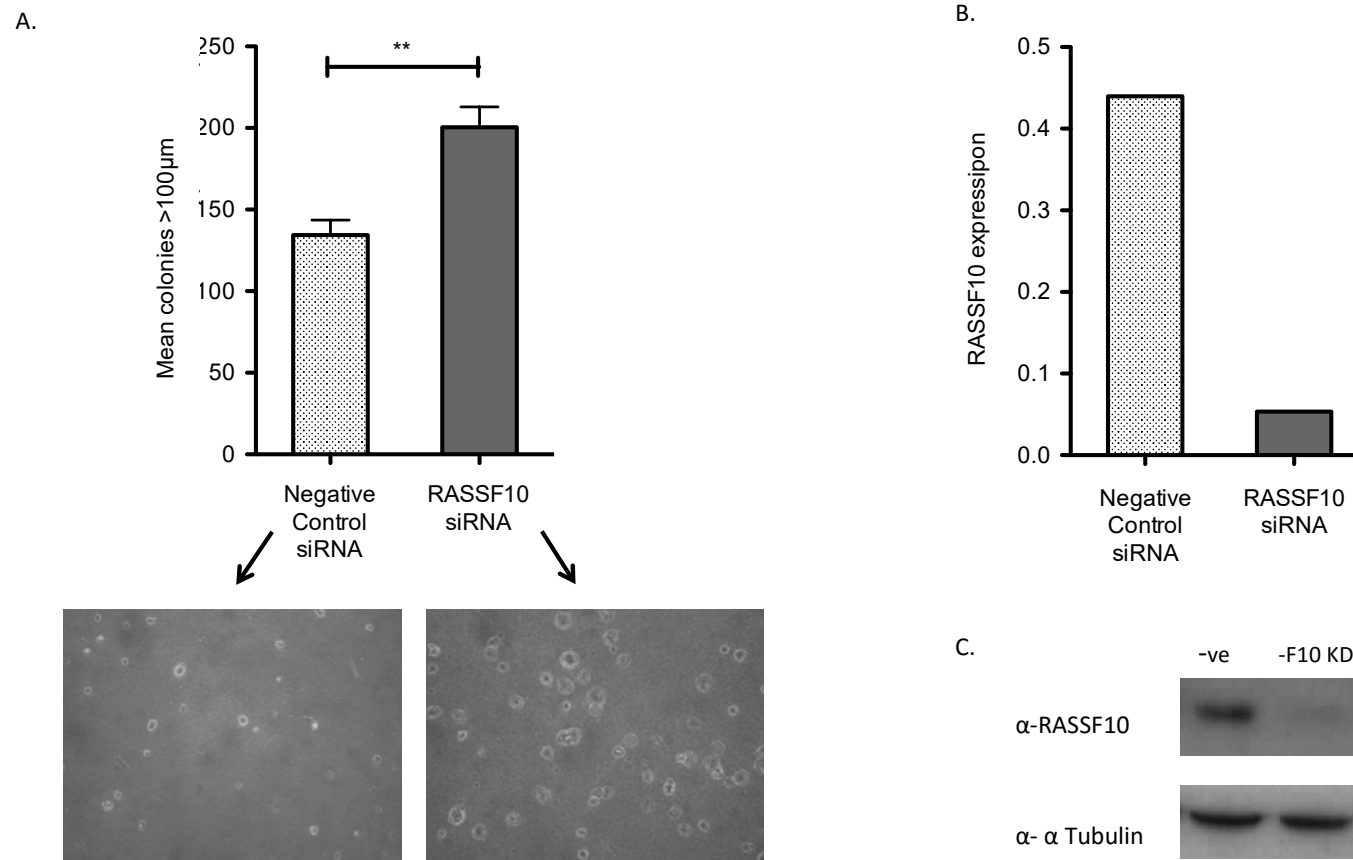
Figure 4-11: RASSF10 effect on cell growth in KTCL45
RASSF10 was knocked down in KTCL45 kidney cancer cell line by siRNA. **A)** Soft agar growth assay shows that knockdown of RASSF10 significantly increased colony formation and size. (t-test P=< 0.0001) Bar chart displays average colony count across repeats (n=3) + SEM**. B)** Quantification of western blot and **C)** Western blot image confirmed siRNA knockdown of RASSF10 correlates with knockdown of the protein. α-tubulin was included as loading control.

Knocking down RASSF10 via siRNA resulted in reduced protein expression and a significant increase in cell growth and colony formation in soft agar, suggesting that RASSF10 acts in a similar manner to a TSG, and the presence of the protein restricts and limits cell growth.

4.1.14 Preliminary investigation of RAFFS10 binding partners

RASSF10, is a member of the N-terminal RASSF family. Although N-terminal RASSF proteins have been identified to display TSG activity (V K Hill et al., 2011; Wei et al., 2013), the protein-protein interactions and tumour suppressive mechanisms of action of this subset of RASSF family of proteins remains poorly understood (Underhill-Day et al., 2011b).

The identification that many of the classical RASSF proteins (RASSF 1- RASSF 6) play a pivotal role in the regulation of cell cycle and apoptosis through activation of the Hippo pathway (Fausti et al., 2012). Coupled with, previous studies identifying that RASSF10 localises to the centrosomes during mitosis (V K Hill et al., 2011). Led us to formulate the hypothesise that RASSF10 may too play a role in cell cycle regulation.

We therefore proposed to investigate the protein interactions of RASSF10. By using immunoprecipitation (IP) to isolate the proteins interacting with RASSF10 at different stages in the cell cycle and mass spectrometry (MS) to identify the interacting proteins. We hope to identify if RASSF10 interacts with known cell cycle proteins, and at which stage in the cell cycle do these interactions occur.

However, in order to associated any protein interactions with specific stage of the cell cycle, cells must first be synchronised so they the majority of cells are at the same stage

in the cell cycle prior to IP and MS. By subjecting growing cells to different treatments or agents, it is possible to synchronise cell cycle so that the majority of cells are in the same phase. In this study (conducted by Dr. D. Matallanas and team (Systems Biology Ireland, University College Dublin)) stable HEK293 cell line overexpressing RASSF10- Flag were cultured under the following conditions: growing (under normal conditions, control group), serum starved, arresting cell cycle at G2/M phase ,100 ng/ml Nocodazole (noco x1) arresting cell cycle at G2/M phase and 2 mM Thymidine (T1) arresting cell cycle at G1/s phase) (section2.7.6) ((Rosner et al., 2013)). Following synchronisation, a co-IP specific for the Flag tag was performed, this would pull down and all proteins currently bound to the overexpressed RASSF10-Flag protein. Bound proteins were then identified and quantified (label free quantification (LFQ) by MS.

Dr. Matallanas and myself conducted joint analysis of the output data (section2.7.6) , firstly the specificity of the original immunoprecipitation was confirmed, this ensured that the IP was specific for RASSF10-Flag, and as a check for contamination and was achieved analysing an untransfected control IP MS output, where the bait protein (RASSF10) should not be observed. Secondly the quality of the IPs and the MS was assessed. This was achieved by comparing the abundance of the bait protein RASSF10, calculated as label free quantification intensity ((LFQ) a quantitative value relating to the abundance of a protein, determined by the MS signal intensity), across all growing conditions containing the RASSF10-Flag transfection, the LFQ for RASSF10 should be present at similar levels across all conditions. We observed that the bait protein RASSF10 was present in all conditions, however it was noted to be lower in the normal

214

growing condition compared to the cell cycle synchronised conditions. It is therefore, imperative that the mass spectrometry be repeated, before definitive conclusions can be drawn from the data. Despite this set back, it is still possible to obtain a preliminary indication into potential RASSF10 binding partners, at different cell cycle stages, although it is imperative that validation and confirmation is conducted.

To identify RASSF10 binding proteins, firstly the LFQ of each protein identified for each condition was compared to the negative control. Proteins with a LFQ fold difference of >1 and a P value <0.001 in the condition compared to the negative control were classified as specific RASSF10 binding proteins. Samples that did not meet these criteria were excluded and deemed to be contaminants. The next stage was to compare between the growing conditions, here the criteria was set to a fold difference in LFQ intensity of >2 fold and a P value of <0.05 (FDR corrected student t-test). When comparing the cell cycle synchronisation growing condition to that of standard growth, MS identified 199 proteins (42 + 27 ribosomal proteins increased in starved, 22 + 17 ribosomal proteins increased in Noco-x1 and 53 + 41 ribosomal proteins increased in Thy x1) with a significant increase in LFQ (FDR<0.05), and thus are candidate RASSF10 binding partners (Table 4-1 to Table 4.3).

Increased in Starved vs Growth					
Protein	**LFQ ratio**	**FDR**	**Protein**	**LFQ ratio**	**FDR**
PGAM5	2.754	8.08E-09	MTMR14	4.061	2.99E-03
EIF3F	3.861	2.92E-08	SRSF3	2.349	2.99E-03
ACAP2	2.185	6.69E-07	ACIN1	3.687	3.13E-03
RASSF10	3.619	6.67E-06	SF3B2	2.189	3.89E-03
BCLAF1	3.196	1.25E-05	INTS7	3.197	3.89E-03
THRAP3	3.478	1.57E-05	OTUD4	3.248	4.85E-03
EIF3E	3.148	1.62E-05	HUWE1	2.068	0.01
EIF3H	2.608	3.85E-05	SAFB	4.248	0.01
TP53BP2	**4.006**	**7.58E-05**	TRA2B	4.940	0.01
OPA1	15.425	1.66E-04	HIST2H2BF; HIST1H2BK;		
RBM10	3.921	2.39E-04	HIST1H2BD; HIST1H2BC;		
SRSF6	2.501	3.26E-04	HIST1H2BH; HIST1H2BN;		0.01
EIF4B	4.322	3.26E-04	HIST1H2BM; HIST1H2BL;		
STK38	2.141	3.65E-04	H2BFS	5.217	
GTPBP10	7.239	3.65E-04	KIF4A;KIF4B	2.147	0.01
EIF4G2	2.189	4.98E-04	RBMX;RBMXL1	4.649	0.01
TRA2A	5.674	5.13E-04	NCOA5	4.240	0.01
EIF3EIP; EIF3L	2.780	8.48E-04	RALY	2.733	0.01
			PRSS1;PRSS2;TRY6	17.551	0.01
GATAD2A	2.278	1.65E-03	HIST3H2BB	17.983	0.02
SPIN3	2.953	1.65E-03	UHRF1	2.514	0.03
TET2	3.035	2.25E-03	ANXA1	5.858	0.05
ERH	2.363	2.33E-03	+ 27 ribosomal proteins		
SRSF1	2.014	2.53E-03			
MATR3	3.007	2.53E-03			

Table 4-1: Protein list of candidate RASSF10 binding partners significantly increased in G1/S cell cycle stage compared to normal growth.
Table displays RASSF10 binding proteins identified by mass spectrometry, protein name, the fold difference of RASSF10 binding proteins between the two growth conditions (LFQ ratio) calculated as LFQ starved/LFQ growth. Proteins identified in other growing conditions are highlighted in yellow, with the most interesting in bold and RASSF10 (Bait protein) in green.

Increased in Noco vs Growth		
Protein	LFQ ratio	FDR
MATR3	4.240	2.49E-05
TRA2B	6.307	2.96E-05
ACIN1	5.214	1.18E-04
TRMT1	2.823	1.21E-04
RBMX; RBMXL1	11.473	2.06E-04
RALY	4.509	3.42E-04
TRA2A	7.365	4.18E-04
ELAVL1	2.285	7.58E-04
SRSF3	2.272	2.47E-03
SF3A1	2.286	3.78E-03
SAFB	12.644	3.72E-03
INTS7	3.318	7.14E-03
SRSF6	2.089	0.01
GATAD2A	2.023	0.01
PRSS1; PRSS2; TRY6	89.385	0.02
OPA1	20.976	0.02
HIST3H2BB	4.223	0.03
AK1	3.282	0.03
SF3B2	2.289	0.04
NCOA5	4.369	0.04
ARPC1A	2.719	0.04
HUWE1	2.433	0.05
+ 17 ribosomal proteins		

Table 4-2: Protein list of candidate RASSF10 binding partners significantly increased in G2/GM cell cycle stage compared to normal growth.
Table displays RASSF10 binding proteins identified by mass spectrometry, protein name, the fold difference of RASSF10 binding proteins between the two growth conditions (LFQ ratio) calculated as LFQ Nocodazole (noco x1)/LFQ growth. Proteins identified in other growing conditions are highlighted in yellow, with the most interesting in bold and RASSF10 (Bait protein) in green.

Increased in T1x vs Growth					
Protein	LFQ ratio	FDR	Protein	LFQ ratio	FDR
EIF4G2	0.40140861	4.00E-04	RASSF10	0.20732901	0.02
PDCD4	0.11345595	4.00E-04	AARS2	1.2996E-06	0.02
USP7	0.46819728	4.00E-04	EIF3A	0.34532816	0.02
DDX28	0.05377929	2.10E-03	**TP53BP2**	0.18589489	**0.02**
ERLIN2	0.37590089	2.10E-03	WDR26	0.33352871	0.02
DHX40	0.32353075	2.42E-03	ARHGAP17; ARHGAP44	0.48584508	0.02
SLC25A10	0.28908757	2.80E-03			
GTPBP10	0.37763337	2.80E-03	GALK1	0.38650469	0.02
WNK1	0.14970391	3.25E-03	LRCH2	0.15418427	0.02
EIF3B	0.40932966	3.96E-03	CECR5	0.47582966	0.02
RPS19BP1	0.23810026	3.96E-03	EXOSC6	0.49289435	0.02
NKRF	0.31887929	4.26E-03	NDUFB8	0.34568703	0.02
USP9X	0.38416164	4.26E-03	EIF3G	0.39995842	0.02
EIF3K	0.456397	4.61E-03	EIF3J	0.32779262	0.02
SMC2	0.44158279	4.96E-03	ICT1	0.18973703	0.02
EIF3F	0.42294883	0.01	GTF3C5	0.47772165	0.02
ARMC6	0.19216297	0.01	DBT	0.3798879	0.03
C8orf33	0.49243977	0.01	TBC1D4	0.28415098	0.03
WRNIP1	0.10658945	0.01	EIF3I	0.47334517	0.03
SMEK1	0.48817113	0.01	NDUFS6	0.34937595	0.03
ERC1	0.2797408	0.01	HEATR3	0.47393869	0.03
VANGL1	2.8185E-06	0.01	NDUFS7	0.43002809	0.03
UBE2E3; UBE2E2; UBE2E1	0.34831134	0.01	PDAP1	0.49591734	0.03
			YTHDC2	0.39746368	0.04
GPKOW	0.46609774	0.01	GADD45GIP1	0.24798912	0.04
DIAPH1	0.42005677	0.01	PELP1	0.14102917	0.04
PRPS2	0.47669961	0.01	HIBCH	0.31917763	0.05
NCAPH	0.47483477	0.02	+ 41 ribosomal proteins		

Table 4-3: Protein list of candidate RASSF10 binding partners significantly increased in G0/G1 cell cycle stage compared to normal growth

Table displays RASSF10 binding proteins identified by mass spectrometry, protein name, the fold difference of RASSF10 binding proteins between the two growth conditions (LFQ ratio) calculated as LFQ Thymidine (T1)/LFQ growth. Proteins identified in other growing conditions are highlighted in yellow, with the most interesting in bold and RASSF10 (Bait protein) in green.

String-db Kegg pathway analysis identified all three synchronisation conditions to be significantly associated with the ribosomal network (FDR: P = 3.560e-21 starved; P = 1.59e-9 Noco x1; P= p2.339e-23) or the splicosome network (FDR: P = 2.459e-8 starved; P = 1.59e-9 Noco x1; P = 7.179e-4. Other statistically significant pathways and networks that were highlighted, were associated with DNA conformational and chromatin assembly change (FDR P = 1.798e-13 starved), thyroid hormone signalling pathway (FCR P= 2.073-1 Thy x1, mitochondrial electron transport and ATP synthesis coupled electron transport (P=<1.97e-9 Thy x1) (Figure 4-12 -Figure 4-14).

Three proteins (excluding RASSF10) were identified in two cell synchronisation conditions (starved (cell cycle stage G0/G1) and thy x1 (cell cycle stage G1/S) EIF3F, EIF4G2 and TP53BP2 (tumour protein 53 binding protein 2; ASPP2). EIF3F and EIF4G2 were associated with RNA transport network (Kegg FDR P= p 7.38e-4 starved and P= p 8.099e-4 Thy x1) (Table 4-1 and Table 4-3). TP53BP2 is a particularly interesting protein due to its association with P53, a protein previously identified as a TSG involved in many key regulatory pathways including multiple cell cycle check points, DNA repair and apoptosis (Giono and Manfredi, 2006). The LFQ for TP53BP2 is shown in Figure 4-15 and although not significant after FDR correction, it was observed that the LFQ for TP53BP2 within Noco x1 growing conditions was 4.6 fold greater than the recorded LFQ for growing conditions. Combined with the identification within the starved and thyx1 growing conditions, it is suggested that TP53BP is an interesting binding partner of RASSF10 during the cell cycle. Additionally P53 is frequently mutated in many cancers (Muller and Vousden, 2013; Rivlin et al., 2011). Targeted hindsight analysis of the MS data identified TP53 within the original output file, however it was excluded during

219

analysis as one of the control replicates presented with a high LFQ (Figure 4-16a). If the anomaly was to be excluded, TP53 appeared to show a significant increase in cell cycle associated conditions when compared to growing (T-Test: P>0.001 starved; P=0.036 Noco x1, Thy1 P=0.005) (Figure 4-16b). If a true observation, it maybe proposed that RASSF10 acts as scaffold protein between TP53 and TP53BP2. However, this proteomics data is from a preliminary study only and ideally requires optimising and repeating. Additional investigation and validation of RASSF10 binding partners is also required and discussed below.
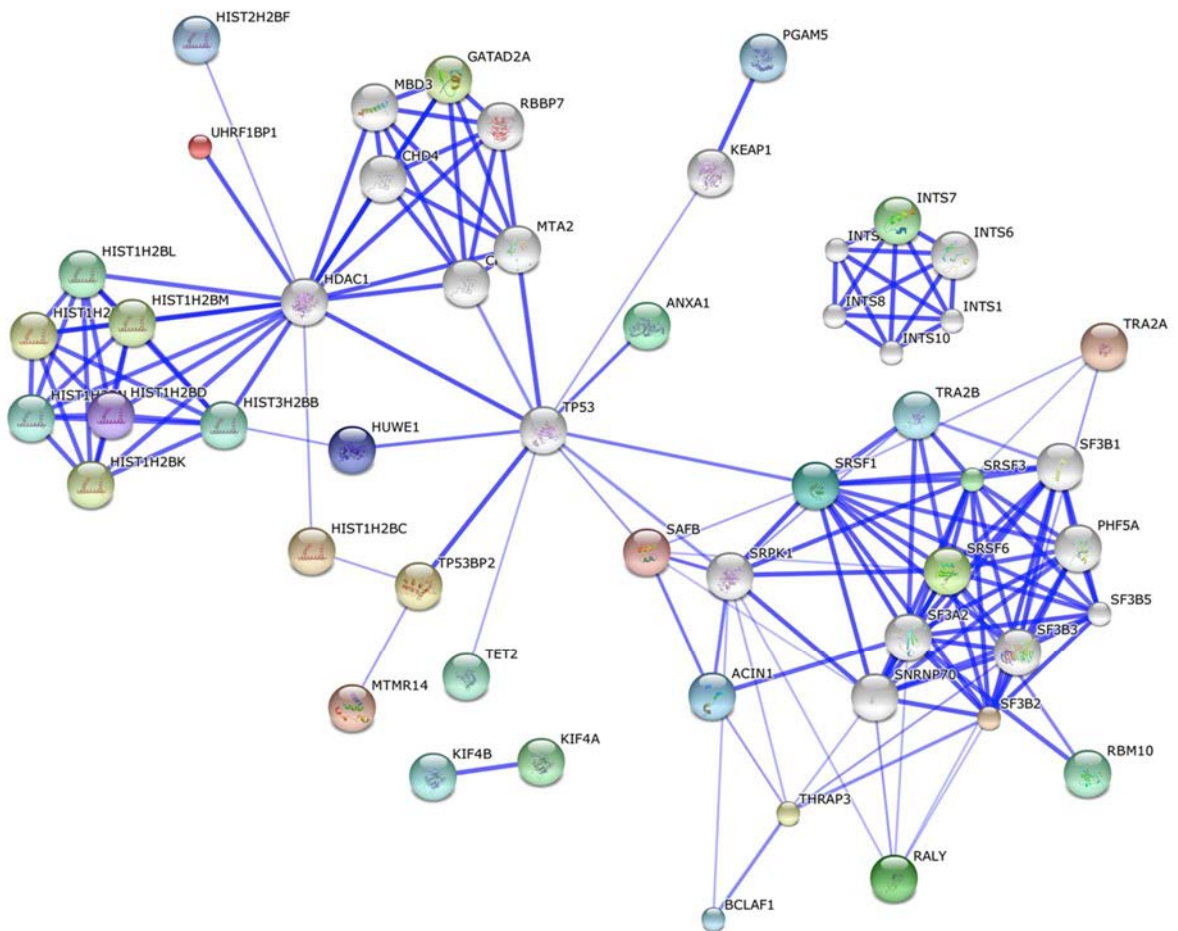
Figure 4-12: String db network analysis: Starved v growth
String-db prediction networks for proteins significantly more abundant in starved compared to normal growth. Blue lines indicate confidence of interaction, all interactions shown are more than median confidence classified as a score >0.4. White nodes are additional predicted connecting nodes and are not present in predicted RASSF10 binding protein list. Ribosomal associated nodes were excluded from string-db image to allow visualisation of other protein networks and interactions.
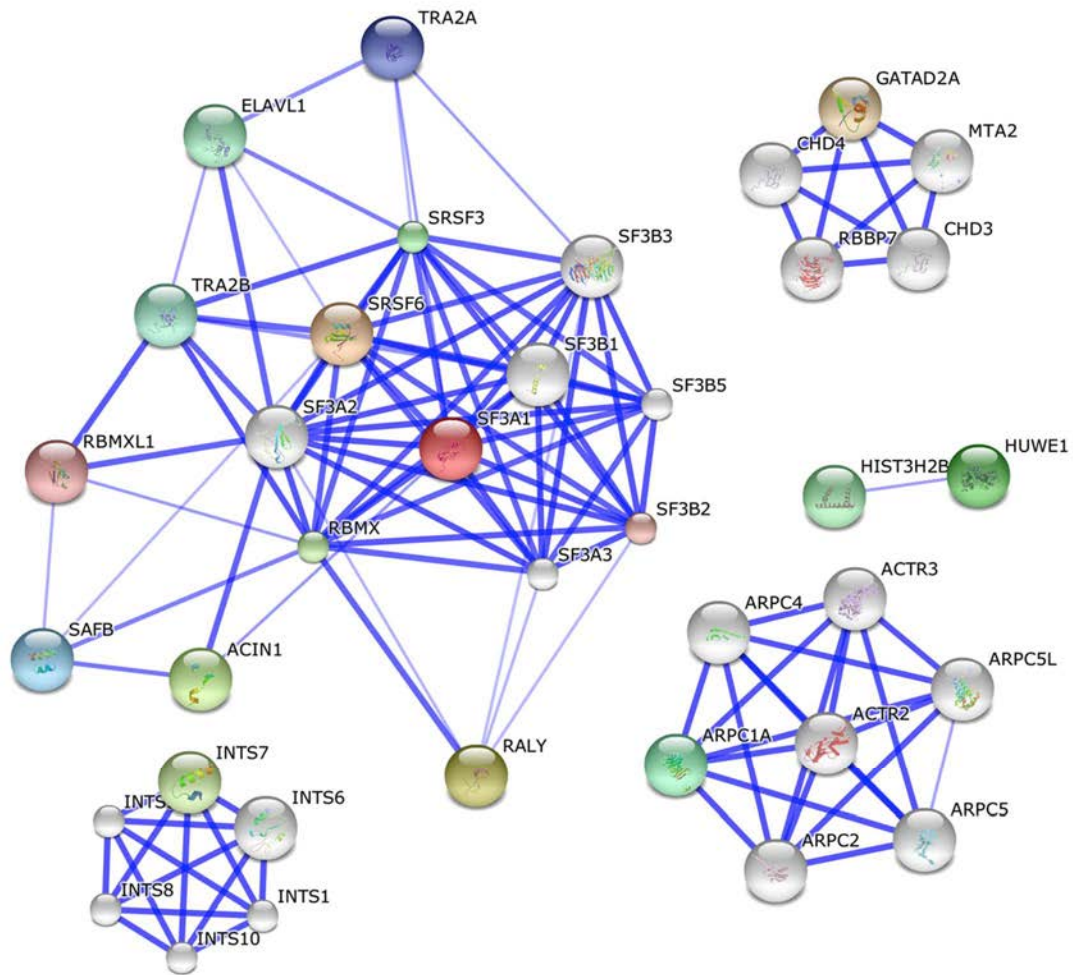
Figure 4-13: String db network analysis: Noco x1 v growth
String-db prediction networks for proteins significantly more abundant in Noco x1 compared to normal growth. Blue lines indicate confidence of interaction, all interactions shown are more than median confidence classified as a score >0.4. White nodes are additional predicted connecting nodes and are not present in predicted RASSF10 binding protein list. Ribosomal associated nodes were excluded from string-db image to allow visualisation of other protein networks and interactions.
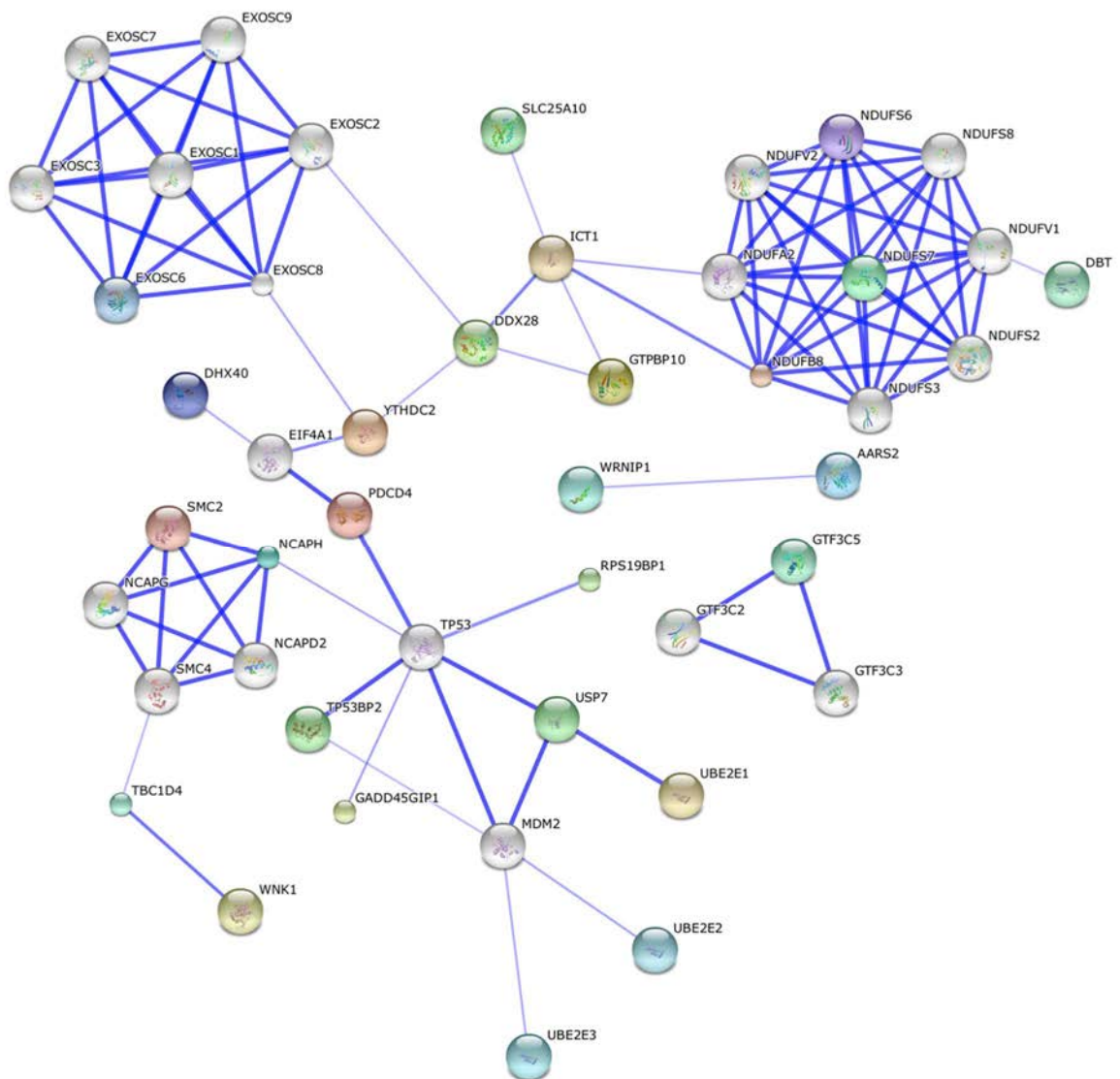
Figure 4-14: String db network analysis: Thy x1 v growth
String-db prediction networks for proteins significantly more abundant in Thy x1 compared to normal growth. Blue lines indicate confidence of interaction, all interactions shown are more than median confidence classified as a score >0.4. White nodes are additional predicted connecting nodes and are not present in predicted RASSF10 binding protein list. Ribosomal associated nodes were excluded from string-db image to allow visualisation of other protein networks and interactions.
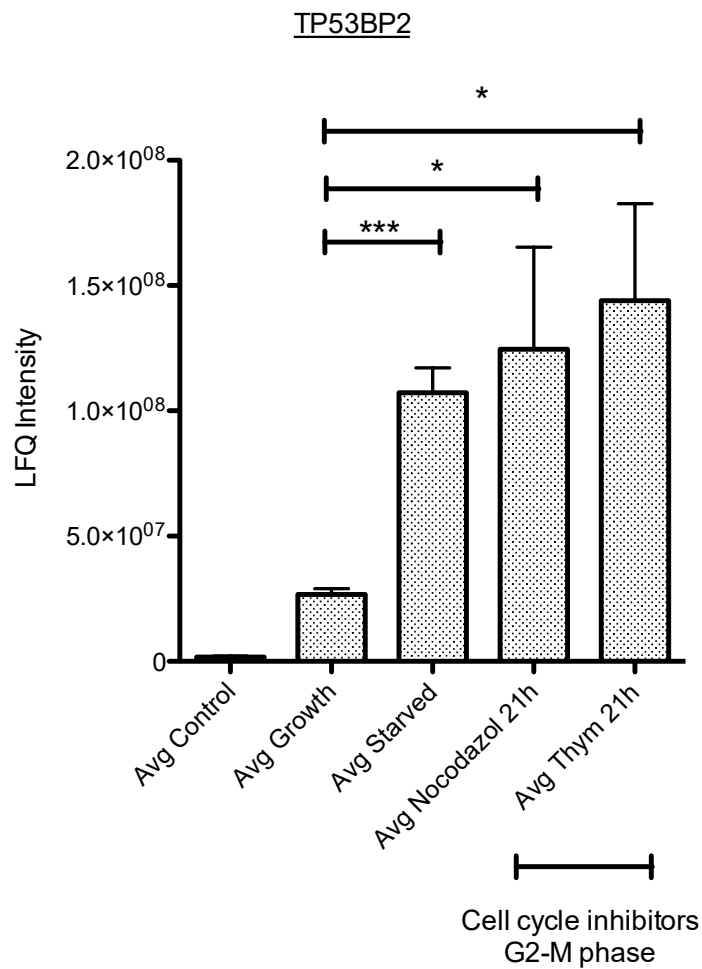
Figure 4-15: Mass Spectrometry reading for TP53BP2
Average LFQ recorded for the TB53BP within each growing condition + SEM. Student T-Test identifies significance between group * P>0.05, and *** P>0.001
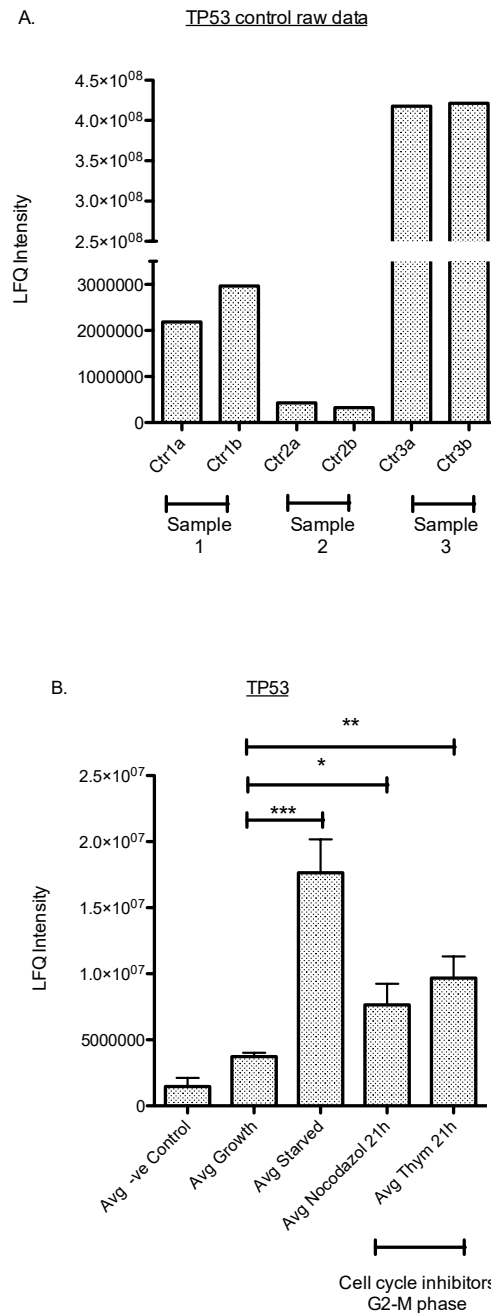
Figure 4-16: Mass Spectrometry reading for TP53
A) Raw LFQ recorded for each replicate of the negative control. Sample 3 was identified to have a much higher LFQ reading for TP53 than the other two repeats. This observation is indicative of contamination within sample 3 **B)** Shows the average LFQ recorded for TP53 within each growing condition, (abnormal repeat within negative control excluded). Student T-Test identifies significance between group * P>0.05, and *** P>0.001

4.2 <u>Discussion</u>

sRCC is a rare and highly aggressive advanced staged RCC, associated clinically with a high rate of metastasises and poor prognosis. However, despite the aggressive nature of the tumour, little research has been conducted into the epigenetic profile of sRCC. Here we have undertaken quantitative Methylight analysis of frequently methylated genes *RASSF1a, RASSF2, RASSF6* and *RASSF10* with the aim of identifying differential DNA methylation, which may provide an insight into the progression and development of sRCC.

4.2.1 Methylation analysis

Within this study we identified *RASSF1A, RASSF2* and *RASSF10* to be frequently methylated in 93.5%, 25.8% and 48.3% sRCC samples respectively.

*RASSF1*A has been frequently reported as being hypermethylated in a cancer specific fashion in a multitude of cancers including, but not limited to, pRCC, ccRCC, chRCC and oncocytoma (Costa et al., 2007; Ellinger et al., 2011b; Morrissey et al., 2001) pheochromocytoma (Richter et al., 2015), lung cancers (Grawenda and O'Neill, 2015), brain gliomas (Hesson et al., 2004) breast cancer where *RASSF1a* methylation is associated with advance stage (Hagrass et al., 2013) and hepatocellular carcinoma (Dong et al., 2015; Xu et al., 2013). Despite this plethora of cancer specific methylation, we identified methylation of the *RASSF1A* promoter region in 100% (4/4) of the normal non-diseased kidney, and thus deemed *RASSF1A* methylation to be a general event and not specific to sRCC. Furthermore quantitative-MSP (q-MSP) upon hepatocellular

carcinoma identified a subset of normal tissue (5%) to present with high methylation of *RASSF1a* with a PMR>50 (Xu et al., 2013). However, it should be noted that q-MSP is not as sensitive or as specific as Methylight and thus the PMR values are not directly comparable. A further study utilising the sensitive Methylight technique, aimed to assess the possibility of analysing *RASSF1a* methylation within the serum of patients with hepatocellular carcinoma, in the hope that *RASSF1A* methylation could be used as a non-invasive biomarker for disease diagnosis. This study demonstrated that methylation could be observed in patient serum and correlated with progression and prognosis (Dong et al., 2015). These findings demonstrate that *RASSF1a* methylation can be identified in many different tissues type and although more frequently observed as methylated in cancers, it may not be solely cancer specific.

RASSF2 was identified to be differentially methylated in 25% of sRCC samples when compared to the normal kidney, however the level of methylation was not identified to be statistically significant. Further comparison with clinical data identified no trend between methylation status and survival, age, gender, size of tumour or occurrence of metastasis thus *RASSF2* was not investigated further.

RASSF10 on the other hand was identified as the most interesting candidate of the genes tested, with frequent methylation observed in 48.3% of sRCC samples and an average PMR of 17%. Furthermore we identified higher methylation levels (average PMR 27% in UICC stage IV ccRCC compared to the PMR average of 6% in UICC stage I) and occurrence in the more advanced stages of ccRCC without sRCC components, suggesting that *RASSF10* methylation is a late stage event that may be involved in the transformation of

ccRCC histology to that of sRCC. However, it is should be noted that the number of samples we were able to investigate for each stage of ccRCC was limited and that frequency and average PMR is likely to be lower than we observed in this study. Additionally we observed a trend that *RASSF10* methylation correlated with increased occurrence of metastasis (P >0.05). This observation requires further investigation and confirmation in an independent cohort. Although we identified no somatic mutations, or LOH, we did show that *RASSF10* methylation reduced mRNA expression within kidney cancer cell lines consistent with the current view on the effect of DNA methylation. Additional soft agar assays were also conducted, in which we knocked down RASSF10 expression via siRNA, this resulted in an increase in cell growth and colony formation. This result coincided with two independent studies and provides further evidence that *RASSF10* is a tumour suppressor gene, which when lost, results in increased cell growth (V K Hill et al., 2011; Richter et al., 2012; Wei et al., 2013)

Many other reports have observed frequent methylation of *RASSF10* in many different cancers such as, childhood leukaemia's (88%) (Hesson et al., 2009), gastric cancer (75%) (Wei et al., 2013) (although a further study of 300 gastric cancer samples put the frequency of methylation much lower at 30% but identifies a significant decrease in cancer-specific survival in patients presenting with *RASSF10* methylation (Deng et al., 2014)), adrenal medulla tumour pheochromocytoma (73%) (Richter et al., 2015) and Sarcoma (17%) (Richter et al., 2012). Helmbold et al., 2012 identified *RASSF10* promoter methylation in 68% of primary malignant melanoma. However, when analysing the secondary tumours resulting form malignant melanoma metastasis (including skin

lymph node and brain) the frequency of *RASSF10* methylation was identified in 91% of the secondary cancers. The differentiation in occurrence of *RASSF10* methylation between primary and secondary tumours, further strengthens our hypothesis that *RASSF10* has a role in metastasis. Hill et al., 2011 studied *RASSF10* methylation in relation to WHO giloma stage. Here they identified *RASSF10* methylation to be absent in grade I giloma, but present in grade II, III and IV (60%, 80% and 65% respectively) suggesting that *RASSF10* methylation occurs fairly early on in tumour development but is also associated with the more aggressive form of the cancer. Li et al., 2014 identified a significant increase in the observation of *RASSF10* methylation in gastric cancer patients with lymph node metastasis compared to patients without lymph node metastasis. Finally, Guo et al., 2015 studied the methylation status of *RASSF10* within 89 primary colorectal cancer samples and identified 60.7% of samples to be positive for *RASSF10* methylation, with no methylation identified within the normal colorectal mucosa. Moreover, analysis of the clinical data identified *RASSF10* methylation to be significantly associated with the more advanced stages of colorectal cancer, it was also identified that *RASSF10* methylation was significantly associated with metastasis to the lymph nodes.

These findings correlates with our hypothesise that *RASSF10* methylation is associated with advanced and aggressive tumours such as sRCC.

4.2.1.1 *Limitations and future work*

The biggest limitation of this study was regarding the histopathology of the sRCC samples. Cancers are very heterogeneous tissues by nature due to clonal evolution in combination with genomic instability and high mutation rates, levels of genetic diversity

can be extremely high, within a single tumour (Gerlinger et al. 2014). sRCC derived from ccRCC, are by definition heterogeneous. This heterogenetic diversity has led to the visually distinct histology, the sarcomatoid features. Furthermore, the extent of sarcomatoid features within a tumour can vary drastically, ranging from 1% to 100% of the tumour histology (Kim et al., 2015). For our cohort, DNA was extracted from the whole tumour by our collaborator Dr M Yao, therefore our analysis was conducted on a mix of sarcomatoid and ccRCC DNA. These heterogeneous samples being investigated, coupled with no clinical data regarding the extent or percentage of sarcomatoid features within the tumours makes the interpretation of this data challenging and limiting. For example, we in this study we identified a PMR range for *RASSF10* of 0 to 116.4. Therefore if *RASSF10* methylation is truly as we hypothesise and associated with the later stages of advanced and aggressive ccRCC, then one would expect the highest levels of *RASSF10* methylation to correlate with higher percentage of sRCC subtype, however, we are unable to draw such conclusions in this study. Furthermore, no data was provided regarding the extent of sarcomatoid features in the ccRCC cohort, therefore objective comparisons between methylation status and stage in this tumour type may be flawed. To overcome this limitation, it would be preferable to obtain sRCC samples, from which the sarcomatoid features have been microdissected from the tumour, not only would this provided a 'pure' histology to analyse, but would also permit the direct comparison of sarcomatoid to the matched ccRCC allowing for investigations into differential methylation of the two histologies (Gillespie et al., 2001; Yoshihara et al., 2013)

Another major limitation, was the number of samples investigated, however due to the rarity of sRCC it was not possible to source a larger cohort, additionally our cohort size for ccRCC examination was limited to the samples for which we had clinical data regarding their stage. Additionally we had no access to sRCC patient RNA or sRCC cells lines so were unable to assess the effect of *RASSF10* methylation in a sRCC model, although the cell lines utilised were derived for a ccRCC cancer, the same tumour type from which the sRCC we studied were derived. A larger cohort would also help to clarify if the observation between *RASSF10* methylation and metastasises is a true association or whether it was just a coincidental observation.

Additionally it should be noted that *RASSF10* methylation was not identified within a ccRCC cohort of TCGA methylation. We propose that this discrepancy in findings may be due to the TCGA 450K characterising the methylation status of individual CpG dinucleotides which may be infrequently methylated in the less aggressive form of cancer, furthermore as shown in figure 4-6 none of the CpG probes mapped to the region investigated by single colony sequencing or Methylight.

Limitations with the Methylight technology are similar to that with standard PCR based methylation studies, in which only a small fragment of the DNA is analysed for methylation and relies on the general assumption that the DNA methylation pattern identified within a small region is indicative of CpG methylation in the surrounding area (Bibikova et al., 2011; Eckhardt et al., 2006). For Methylight the region investigated is usually smaller than MSP, with primers designed to amplify a product length of between

50- 200bp as longer products can affect the efficiency of the fluorescence of the probes (Trinh et al., 2001).

Furthermore due to the sensitivity of the array it can be highly susceptible to experimental error such as variation within pipetting volume or contamination.

However, despite these limitations, bisulfite sequencing confirmed that the PMR was representative of the surrounding area. Furthermore, by conducting each sample in duplicate on the plate and each plate in triplicate, the risk of outliers was reduced within the data. On the whole we found the array to be reliable, consistent and representative of the DNA methylation status.

Within our study we observed PMR values of >100 for some samples when testing for the methylation status of *RASSF1a* and *RASSF10*. This had previously been attributed to incomplete methylation of the intro-methylated control at that particular region (Selamat et al., 2011; Xu et al., 2013). However no further mention as to the effect of this observation was reported and it is assumed that the sample is fully methylated at that given region.

Future work would involve the further characterisation and investigation of the role of RASSF10 in tumourigenesis. Also we should conduct a second soft agar assay using a different siRNA targeted for *RASSF10*, in order to validate and confirm that the result observed, that loss of *RASSF10* results in increased cell growth was not the result of off target effects of the siRNA transfection. Additionally, further functional studies such as scratch migration assays should be under taken to investigate the association identified

between *RASSF10* and metastasis, and to identify if RASSF10 knockdown effects cell migration.

Further work on the sRCC samples would be to investigate the less frequently methylated members of the RASSF family, using this sensitive technique to identify if there is any further differential methylation patterns. For example *RASSF3* was identified to be frequently methylated in pituitary tumours although no observation between methylation state and grade was observed (Peng et al., 2013), although other studies have not identified any further methylation of *RASSF3* in other cancers. *RASSF5A* methylation on the other hand is fairly common. Several studies on cancers including thyroid and skin cancer identified RASSF5A methylation at high frequency within the normal, thus it was not deemed cancer specific (Richter et al., 2013; Schagdarsurengin et al., 2010).

Other work would require sourcing a second independent cohort of sRCC derived from ccRCC and confirming methylation frequency and association with tumour stage and metastasis. Additionally if RNA could be sourced then we would be able to assess the effect of *RASSF10* methylation on mRNA expression within primary tumour samples.

4.2.1.2 *RASSF10 proteomics*

At the time of this study limited investigation had been conducted into the pathways or binding partners with which RASSF10 was associated. A study by Richter et al., 2013 investigated the effect of PKA overexpression on RASSF10 expression and concluded that PKA is likely to be an upstream activator of RASSF10. A second study, identified that

RASSF10 overexpression induced apoptosis via inhibition of the Wnt/β-catein signalling pathway (Wei et al., 2013).

Previous studies have identified an association and role for other RASSF family members (predominately RASSF1a, RASSF2, RASSF3, RASSF5 and RASSF6) within the HIPPO pathway, apoptosis and cell cycle regulation (Cooper et al., 2009; Iwasa et al., 2013; Matallanas et al., 2007; Romano et al., 2010; Shivakumar et al., 2002). Another study identified RASSF10 localisation during mitosis to the centrosomes (V K Hill et al., 2011). As a result of this, and due to the high homology observed between members of the RASSF family, our initial investigations focused on the analysis of RASSF10 binding protein within cell cycle checkpoints via MS.

Through a preliminary study, it was identified that TP53BP2 maybe an interesting candidate as a RASSF10 binding protein, with intensity of binding significantly increased in cells that had had their cell cycle synchronised (Starved and thy x1). Noco x1 identified an increase in TP53BP2/RASSF10 binding, although this was not a statistically significant observation.

TP53BP2 (Tumour Protein p53 binding protein 2; aka ASPP2 (Apoptosis-stimulating of p53 protein 2)) as its name suggests, binds the C-terminal of the P53 TSG and is involved in the regulation of apoptosis and cell growth through many complex interactions and is associated with impeding cell cycle progression at G2/M phase (Genecards.org, 2015; Chen et al., 2003). ASPP2 is categorised as a TSG and has widely been investigated as a pro-apoptotic gene, driving p53 mediated apoptosis in response to DNA damage (Z. Wang et al., 2013). Together with its counterpart (ASSP1) it has been identified that

ASSP2 activation by the oncogene RAS is required to enhance the tumour suppressive effects of p53. Thus ASPP2 is an important negative regulator of oncogenic RAS associated growth (Godin-Heymann et al., 2013; Y. Wang et al., 2013). Additionally ASPP2 elicits a tumour suppressor function and regulation of RAS by facilitating RAS induced senescence (Y. Wang et al., 2013; Z. Wang et al., 2013). Normal cells possess multiple safety features in order to protect against tumourigenesis; one particular feature is senescence. This is where a cell permanently exits the cell cycle and can no longer proliferate, this feature is lost in cancer. When an oncogene such as RAS is activated, so too is the senescence pathway of the primary cell and the cell is withdrawn from the cell cycle; a process known as oncogene-induced senescence (Dimauro and David, 2010; Karnoub and Weinberg, 2008). A recent study characterising the N-terminal of APSS2, binding to RAS-GTP and inducing Ras/Raf/MEK/ERK signalling this in turn facilitates activation of RAS-induced senescence independently of P53 activation (Wang et al., 2013).

Other studies have identified additional roles for ASSP2 with regards to cell polarisation, cell junction formation and central nervous system development. Reporting that loss of ASPP2 resulted in shortened cell cycle length (possible due to the cell cycle arrest and check points being missed) in neural progenitor cells (Sottocornola et al., 2010)

Having identified TP53BP2, as an interesting potential binding partner of RASSF10 we went back to analyse the raw MS data to investigate if TP53 was identified and why it did not pass our thresholds. This hindsight analysis identified that one of the negative control samples presented with a high LFQ reading for TP53, resulting in the TP53 not

passing our reliability thresholds of >1 fold difference in RSSF10-FLAG+ cells with a significance of P<0.001. This outlier may have been due to contamination within one sample, however excluding the sample resulted in a n=2, so we were unable to repeat the t-test to confirm if TP53 met our reliability criteria. Although this casts doubt over whether TP53 is a genuine RASSF10 binding partner, we did observe the LFQ for TP53 to also be significantly increased in cell cycle synchronised samples. The association of p53 as a cell cycle regulator, particularly its involvement in G1/S and G2/M cell cycle checkpoints (Giono and Manfredi, 2006; Harris and Levine, 2005), makes this observation highly interesting and may indicate a role for RASSF10 in cell cycle regulation. Moreover the identification of another N-terminal RASSF protein, RASSF8, to form protein: protein interactions with the ASPP proteins and stimulate p53 activation in Drosophila provides further evidence that this interaction observed between RASSF10 and ASPP2 may be genuine and of high interest (Langton et al., 2009).

Other RASSF proteins previously associated with the cell cycle, include RASSF3, RASSF5 and RASSF6, which is predicted to stabilise p53, and augment ubiquitination of the p53 suppresser MDM2, with RASSF6 knockdown resulted in increased p53 degradation (Iwasa et al., 2013). RASSF1 has been associated with promoting cell cycle arrest at the G1/S stage and decreasing cyclin D expression, while other evidence suggests a further role of RASSF1 regulating microtubule polymerisation within mitosis (Donninger et al., 2014; Richter et al., 2009; Shivakumar et al., 2002).

4.2.1.3 *Future work and limitations*

Future work would revolve around validating the MS, by co-immunoprecipitation (co-IP) to confirm that

1) RASSF10 binds to TP53BP2 and potentially TP53.

2) Identify if binding is alerted under different growing conditions.

Investigation should also be conducted to ascertain the role and function of RASSF10. It is widely considered that the RASSF family do not possess catalytic activity but instead act as scaffold proteins mediating protein-protein interactions (Cooper et al., 2009; Donninger et al., 2007; Fausti et al., 2012; Ferrell, 2000), we therefore predict RASSF10 to function in a similar way. To confirm this RASSF10 should be transfected into cells in increasing concentrations and co-IPs conducted. If this hypothesis is true, comparison between bound proteins such as TP53BP2 should possess the strongest signal when RASSF10 expression is optimal.

Additionally, the effect of RASSF10 on cell cycle should also be undertaken. This could be achieved by knocking down RASSF10, or overexpressing RASSF10 in cell lines and analysing the cell cycle via flow cytometry.

Before any of this can be conducted, certain limitations of this study need to be addressed. Firstly, being a preliminary study aimed at ascertaining if RASSF10 had any involvement in cell cycle and interaction with cell cycle regulators, rather than fully characterising RASSF10's association in cell cycle. The samples run on this array were not fully optimised for cell cycle synchronisation nor was synchronisation confirmed via

flow cytometry. Additionally, not all conditions tested within this MS were successful and achieved a n=3, for example an additional cell synchronisation group (Thyx2) failed in one sample so had to be excluded as no statistics could be conducted. We have identified that TP53 may bind to RASSF10, however due to one of the negative control samples presenting with a high LFQ this protein was omitted as contamination. This outlier may be a genuine feature of HEK293 MS, or alternatively may have been a result of contamination; this can only be confirmed by repeating the MS.

These concerns would be omitted following optimisation of cell synchronisation before repeating the MS, and hopefully more light would be shed on the likely binding of RASSF10 to TP53 and TP53BP2.

Furthermore, experience of Dr D. Matallanas and team suggests that the high occurrence of ribosomal proteins identified within our analysis is likely to be a contaminate of the co-IP specific for FLAG undertaken to produce the samples run within this study and unlikely to be biologically relevant. Furthermore Dr D. Matallanas has frequently observed the presences of ribosomal proteins during Flag-co-IPs for other studies and proposes a GFP based co-IP to be cleaner. For our study this would require generation of a RASSF10-GFP construct and generation of a HEK293-RASSF10-GFP stable cell line.

4.2.1.4 *Recent publication*

As previously mentioned, a recent study by Guo et al., 2015 has since shown RASSF10 to be frequently methylated in 60.7% primary colorectal cancers, additionally they identified a significant association between methylation of *RASSF10*, cancer stage and metastasises. They subsequently investigated the function of *RASSF10* and identified that

overexpressing RASSF10 in cell line that did not endogenously express RASSF10, resulted in a significant decrease in colony formation, thus confirming its function as a TSG. Investigation into the cell cycle of these cells, showed that RASSF10 expression increase in the G2/M phase cell cycle phase, further analysis of key markers (cdc-2 and cyclin-B) associated with cell cycle progression from G2 into M were downregulated. Thus Guo et al. concluded RASSF10 induces G2/M cell cycle arrest. They then conducted microarray experiments for gene expression between the cell lines expressing RASSF10 and the vector control. Here they identified expressions of the p53 suppressor protein MDM2 to be significantly decreased when RASSF10 was expressed. Further analysis of the role RASSF10 played on p53 and MDM2 was conducted by subjecting cells to UV DNA damage. Guo et al. showed that in cells with DNA damage, expression of RASSF10 resulted in increased p53, BAX and cleaved caspase-3 expression and reduced mdm2 expression. Knockdown of RASSF10 in cell lines endogenously expressing RASS10 resulted in impaired p53 expression. The final section of this study was to establish the effects of RASSF10 *in-vivo* here it was shown that RASSF10 expression in xenograft mice models resulted in reduced tumour formation.

From this study it was concluded that RASSF10 is a tumour suppressor gene which is frequently lost in late stage cancers and metastasis, furthermore they conclude that RASSF10 is involved in p53 activation and can elicit apoptosis and G2/M cell cycle arrest in colorectal cancer cells.

This study corresponds with our hypothesis that RASSF10 methylation is associated with advanced cancer stage and metastasis, and that RASSF10 is involved in cell cycle regulation, via association with P53.

4.2.1 Final summary

In this study we have identified *RASSF10* to be frequently methylated in sRCC tumours derived from ccRCC. We propose that methylation of *RASSF10* is associated with increased cancer stage and an increased risk of metastasis, an observation that seems to be consistent with reports of *RASSF10* methylation in other cancers. Moreover we identified *RASSF10* methylation has a negative impact on mRNA expression, which is reversed when DNA methylation is removed by 5-Aza treatment. Additionally we concluded that *RASSF10* functions as a TSG, suppressing growth *in-vitro*, with knockdown of *RASSF10* via siRNA resulting in an increase in colony formation. This too is consistent with current literature.

Finally an attempt to understand the proteomics of RASSF10 and its associated binding partners, identified association with TP53BP2 (ASSP2) within cells with synchronised cell cycle. Although further investigation is required to confirm and characterise the binding and role of RASSF10 with TP53BP2, this data does suggest that RASSF10 may be involved in cell cycle regulation via P53. A hypothesis supported by J. Guo et al., 2015.

The work undertaken in this chapter was presented at a poster session at: 10th NCRI Cancer Conference; 2014 November 2-5; Liverpool, UK by Slater, A. et.al.

Chapter Five: **Whole Exome Sequencing (WES) of sporadic oncocytomas**

5.1 Introduction

5.1.1 Next Generation Sequencing

Next generation sequencing (NGS) has provided researchers the tools to compile extensive maps and databases of genetic variation including single nucleotide polymorphisms (SNP), small insertions and deletions, and structural variants (Cooper and Shendure, 2011).

Since 1977 when the original Sanger sequencing protocol was published by Sanger et al., 1977, revolutionising the field of DNA investigation, sequencing technology has significantly improved. Leading to the development and release of high-throughput multi-parallel sequencing platforms termed Next Generation Sequencers (Bayés et al., 2011). NGS is therefore a broad term encompassing multiple sequencing techniques such as, RNA-sequencing, targeted sequencing, bisulfite sequencing, whole exome sequencing (WES) and whole genome sequencing (WGS) (Bell, 2010; Reis-Filho, 2009; Simon and Roychowdhury, 2013).

RNA-sequencing, involves the sequencing of the mRNA to identify variants and translocations within the transcriptome, and can also quantify gene expression levels (Reis-Filho, 2009). RNA sequencing has proved invaluable for the identification of novel gene fusions such as *BCOR-CCNB3* fusion in undifferentiated spindle-cell sarcoma (11% of samples) (Peters et al., 2015), or *MITF-TFE* translocation involved in RNA splicing in translocation RCC, (a rare subtype of kidney cancer (accounting for 15% of RCC in

patients under 40 years of age) that is characterised by translocations involving the transcription factor genes *TFE3* or *TFEB*) (Malouf et al., 2014).

Targeted sequencing uses predesigned or custom designed capture panels to isolate specific genes associated with a phenotype prior to sequencing (Jamuar et al., 2015). Such as the ion ampliSeq cancer hotspot panel (Thermo Fisher Scientific), a panel of 46 critical cancer associated genes to capture and sequence a panel of candidate genes of interest (GOI) (Thermo Fischer Scientific, 2015), or the TruSight gene panel range from Illumina. Where gene panels have been developed to capture and sequence GOI associated with inherited disease, cardiac conditions, cancer, (including specialist panels for solid tumours and myeloid malignancies) and many more (Illumina, 2016).

Targeted sequencing is a method favoured by many laboratories due to being cost effective, covering candidate genes with a deeper read depth than WES and requiring less bioinformatic analysis to interpret and identify somatic variants compared to other techniques such as WES (Jamuar et al., 2015). However, a major limitation is that potentially pathogenic novel variants in genes not targeted by the panel would be missed. Targeted sequencing is also a popular tool within diagnostic laboratories, due to the cost benefit while prior clinical predictions of diagnosis based on patients phenotype, allows for candidate GOI to be selected prior to sequencing (Dietel et al., 2015).

WES, (section 5.1.1.1) uses a capture method targeting virtually all exons, providing a global picture of genetic variation and copy number alterations within exonic regions within the genome. WGS, on the other hand sequences the whole genome, providing

additional information for the intronic, promoter and enhancer regions, in addition to the exonic regions (section 5.1.1.2) (Lelieveld et al., 2015).

Since 2007 the cost of NGS has substantially decreased, (Figure 5-1), exceeding the original trend predicted by Moore's Law. (Moore's law was a trend originally observed within computer science, which states that the processing power of computer chips will double every two years. Resulting in a linear trend line (Waldrop, 2016). The reduced costs has made this technology more accessible for smaller laboratories as well as permitting the analysis and profiling of large cohorts of samples such as those undertaken by the TCGA (Davis et al., 2014; TCGA, 2013, 2008) or NHS England 100,000 genome project (Siva, 2015). The influx of data has permitted the generation of comprehensive databases collating information on genetic variation and single nucleotide polymorphisms (SNP) within a general population (such databases include dbSNP, 1000 Genome project, JSNPs, Integrative Japanese Genome Variation Database and ExAC) and cancer associated variants (COSMID), all of which are key bioinformatic tools when interrogating the large data sets produced by NGS. It is widely regarded that identification and characterisation of genetics variants will provide unprecedented insight into disease progression and prognosis and permit predictions of how individuals may respond and metabolise drugs. Paving the way for personalised medicine (Simon and Roychowdhury, 2013).
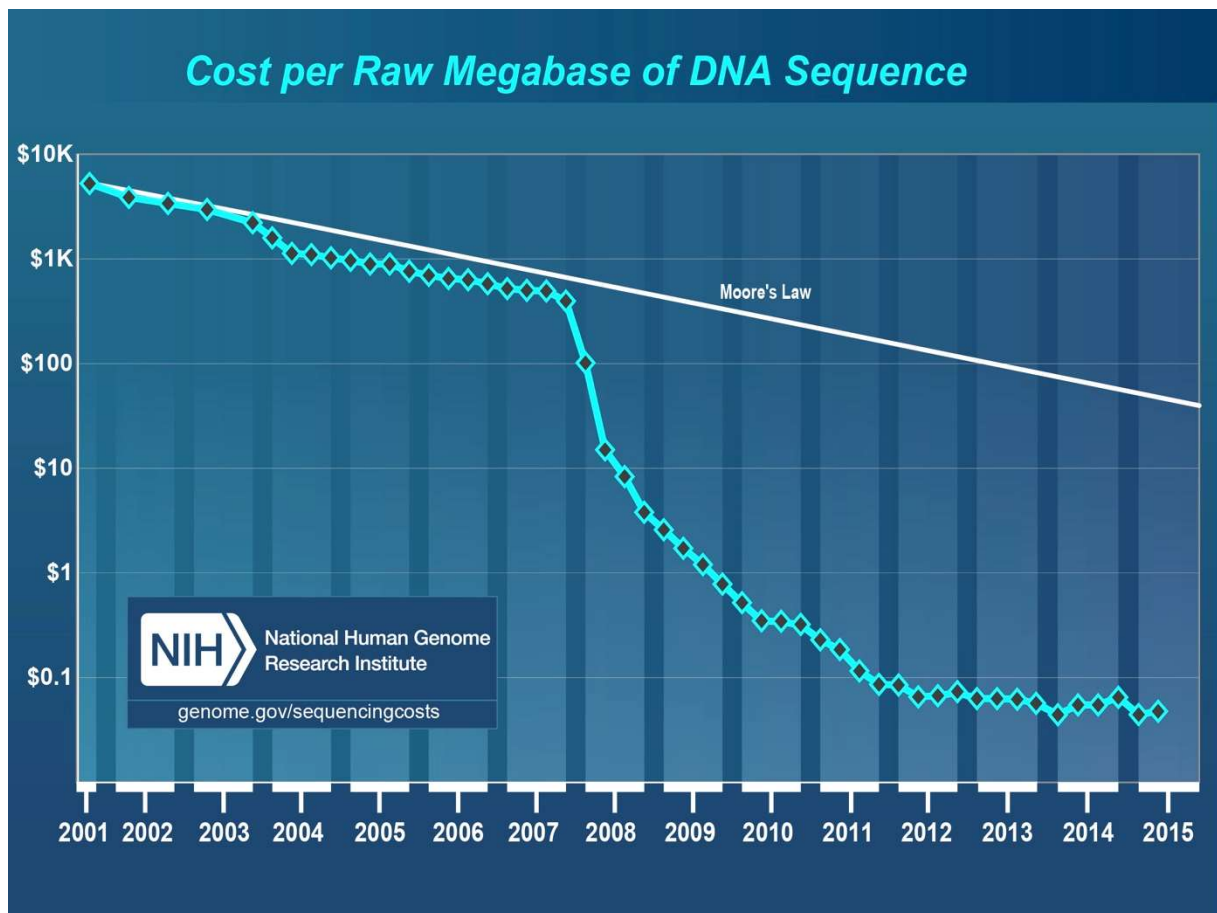
Figure 5-1: Cost to sequence a Mega-base of DNA in comparison to Moore's law
Graphical illustration of the decline in sequencing cost per mega-base of DNA over the last 15 years, in relation to Moore's law. Graph from National Human Genome Research institute Genome.gov. 2015

5.1.1.1 *Whole Exome Sequencing*

As previously mentioned WES involves isolation and sequencing of the exonic regions within the genome. To date several capture and sequencing technologies are commercially available with different methods use to analyse the DNA sequence. Exon capture kits are estimated to isolate 95% of all exons (Lelieveld et al., 2015; Rabbani et al., 2014a). Although other kits are available, in this study exon capture was achieved using the SureSelect exon capture kit (Agilent). This was achieved via PCR amplification of the genomic DNA sample with sequencer specific adaptors, exons were identified via hybridisation to biotinylated RNA library baits specific for exonic regions. These captured exonic regions are then separated from the intronic fragments by streptavidin beads prior to analysis on a NGS sequencer (Figure 5-2). Currently, the market leader in NGS sequencers is Illumina, with sequencers such as the HiSeq2500 being highly popular due the high number of reads per run it can achieve (Sims et al., 2014) . The Illumina sequencers utilises a clonal amplification technique known as bridge amplification, where the template DNA attaches to the array by the binding to adapters, and is amplified by forming 'bridges' with adjacent primers. This generates clusters, each derived from one fragment of DNA, which are then sequenced (Figure 5-3a). The DNA sequence is identified by the incorporation and identification of modified 3' blocked reversible fluorescent nucleotides, cleavage of the fluorescent label permits detection and identification of the incorporated base, and allows the continued extension and interrogation of the template sequence (Figure 5-3b) (Metzker, 2010).
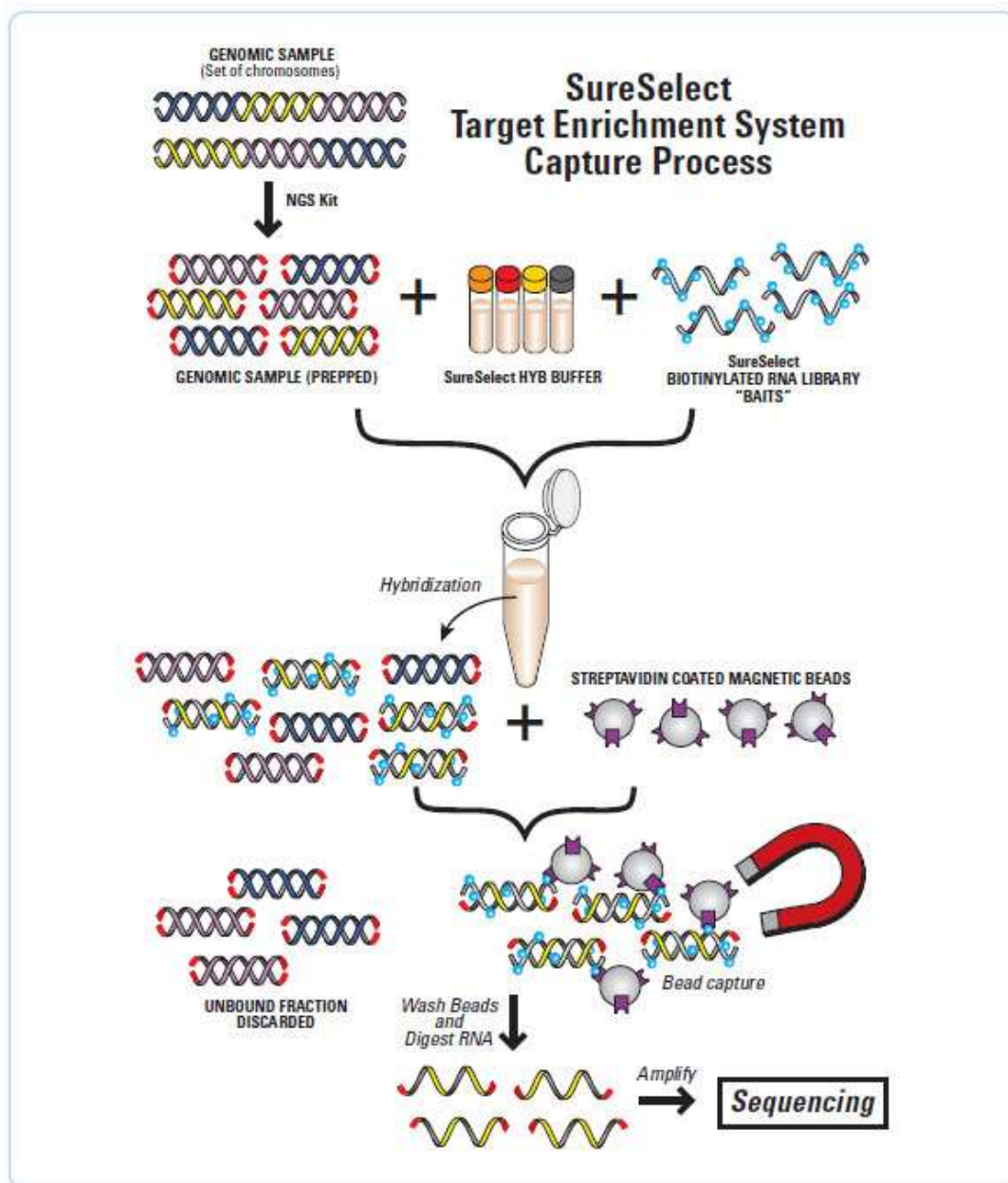
Figure 5-2: Sure Select Exon capture.
Schematic of the method by which the Sure Select exon enrichment kit isolates exonic genomic DNA from a sample. In summary, genomic DNA is fragmented and attached to NGS specific adapters, before hybridisation with exon specific biotinlated RNA. Streptavidin beads permit the extraction of the exonic DNA. DNA is then washed prior to annealing to sequencing array and sequencing. Image from (Genomics.agilent.com, 2015).
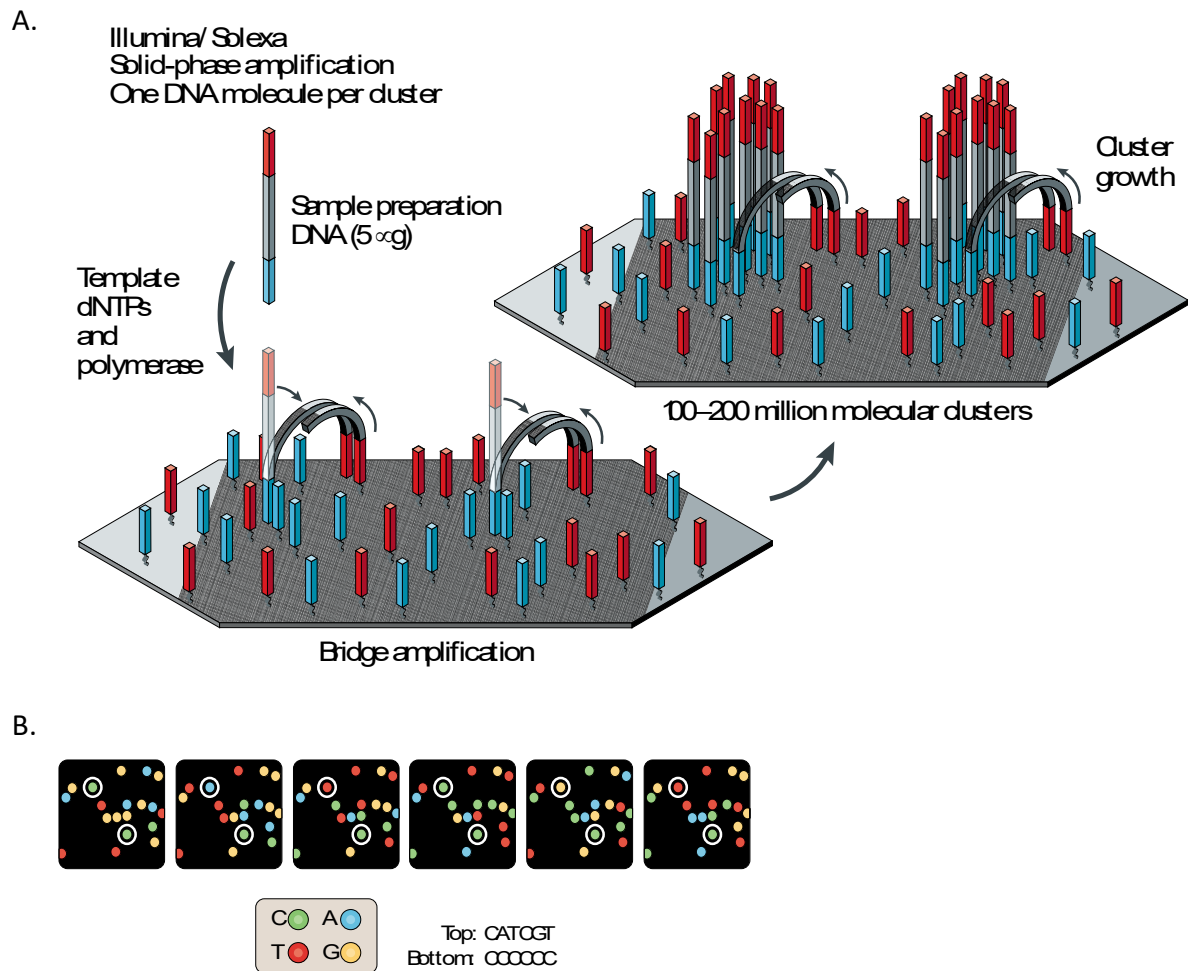
Figure 5-3: Illumina Sequencing technology

Diagram demonstrating the principle behind the Illumina sequencing technology, **A)** Genomic DNA flanked with adapter sequences adhere to complementary adapters on the array, this causes the DNA fragments to form 'bridges' permitting amplification of the DNA and the formation of clusters. **B)** Sequencing of the DNA, involves amplifying the DNA using including fluorescent labelled nucleotides. As the labelled nucleotides are incorporated the fluorescent tag is released and recorded as a colour. Each colour represents an individual cluster. Image adapted from Metzker 2014.

Other technologies such as Roche 454 or ION torrent PGM, immobilise and amplify the template in emulsion on micro-beads. Roche 454 uses fluorescence and the pyrosequencing technique to determine the DNA sequence. While the ION torrent PGM, determines the DNA sequence by measuring a change in the pH as a nucleotide base is incorporated (Metzker, 2010).

The main advantage of WES, is that it is targeted to the 1.5% of the genome that encodes proteins (Koboldt et al., 2012; Meynert et al., 2014), although this is a small percentage of the total genome it is reported that around 85% of disease causing variants are located within the exons (Rabbani et al., 2014a), thus reducing the volume of redundant data produced and the bioinformatic analysis required to mine it (compared to WGS). However there are some limitations, these mainly revolve around the exon capture technology, firstly not all the exons will be captured and secondly of those that are, coverage and representation may not be equal due to PCR bias (described as the unequal amplification of PCR products). This is usually associated with GC rich regions of DNA. These regions usually have increased melting temperature and the ability by which PCR products of these regions can form stable secondary structures, both of which impede PCR amplification (Mamedov et al., 2008). Within WES this PCR bias can result in variation in the number of reads (read depth) of a particular fragment (Ku et al., 2012; Lelieveld et al., 2015).

Additionally the small fragments that are sequenced via the WES technologies (75 - 100bp Illumina; 300bp Roche 454 (Metzker, 2010)) can result in misalignment issues when mapped to the reference genome particularly in highly repetitive regions.

Although WES attempts to combat some of these issues by sequencing at greater depth of multiple short-paired end reads, with the longer reads, providing more accuracy when aligning to the genome. This alone cannot counteract all the issues of alignment especially within repetitive regions that exceed the fragment length (Meynert et al., 2014; Patel et al., 2014; Robinson et al., 2011). Finally, WES is a cheaper technology than WGS, both financially and in terms of labour and computational experience, and is currently the most utilised technique for identifying novel variants in disease (Watson et al., 2013).

5.1.1.2 *Whole Genome Sequencing (WGS)*

Unlike WES, WGS does not require PCR based enrichment of the sample prior to sequencing. This means that the whole genome is mapped and aligned to a reference. The WGS platforms are adapted to permit longer read length allowing more accurate alignment to the reference genome particularly across repetitive regions. Furthermore, WGS platforms are reported to offer improved, more uniform and reliable coverage of coding regions (Lelieveld et al., 2015; Meynert et al., 2014). In addition WGS can provide information on intronic translocations and breakpoint enhancer regions and promoter regions (e.g. the *TERT* promoter see 5.1.2), however the volume of data that is produced makes alignment and interpretation computationally challenging.

Regardless, these challenges are not limiting the use and application of WES and WGS, with studies now being undertaken to identify and characterise unique mutation signatures and frequently mutated pathways such as the focal adhesion pathway, P13k-Akt pathway, Rap1 pathway, and calcium signalling pathways associated with many

different cancers. An investigation that previously would have been limited to candidate genes and expensive and labour intensive to conduct (Alexandrov et al., 2013; Cline et al., 2013; Davis et al., 2014; Neapolitan et al., 2015).

5.1.1.3 *Applications of NGS to investigate human disease*

NGS, has provided unprecedented insights into genetic and structural variations within the general population or diseases. Permitting the identification and characterisation of common single nucleotide polymorphisms (SNPs) present within a general population, or the identification of variants associated with disease phenotype (Cooper & Shendure 2011). One of the major aims and challenge of genetic studies, is identifying pathogenic variants from a vast quantity of non-pathogenic variants and SNPs, and to elucidate the roles they play in disease (Cooper & Shendure 2011).

Within human genetics NGS has been of immense success in the identification of causative variants responsible for inherited Mendelian disease. Especially in the case of autosomal recessive disorders, where homozygous pathogenic variants are often identified homozygous in affected individuals, these can then be mapped to both parents who will be carriers of the disease (Jamuar et al., 2015). One of the first inherited Mendelian disorders from which WES was used to identify the causal variant was Miller syndrome in 2010 (Ng et al., 2010). In the following 2 years, WES had identified the causal variant in over 100 Mendelian disorders (Rabbani et al., 2012). Identification of causative variants of Mendelian disorders, will enable improved diagnostics and prenatal screening and genetic counselling (Jamuar et al., 2015).

In addition, NGS has made many notable contributions to the investigation into the genetics of familial and sporadic cancer. With the identification of driver mutations, and deregulated cancer associated pathways a notable aim of these studies (Rabbani et al., 2014b). By examining the mutations identified between both tumour and normal tissue from an individual, many WES studies have identified somatic cancer-associated variants and alterations. In fact it has been reported that over 150 genes have been identified to possess cancer associated variants (Rabbani et al., 2014b), with most genes identified being classified into 12 signalling pathways, associated with core cellular processes such as metabolism, chromatin modification, transcription, and mRNA processing (Chmielecki and Meyerson, 2014; Kadoch et al., 2013; Vogelstein et al., 2013). However, due to the heterogeneous nature, genomic instability and accumulation of variants with variable penetrance within cancers, identifying driver mutations is a particularly complex challenge (Chmielecki & Meyerson 2014). Despite this challenge, it has not deterred investigations and there have been many advancements in ours understanding of cancer genetics. One such example, by Varela et al 2011, identified the frequently mutated ccRCC associated gene PBRM1 following WES on 7 ccRCC and matched normal DNA (Varela et al. 2011). While WGS studies have been instrumental in the identification and characterisation of recurrent somatic mutations and structural rearrangements within the cancer genome (Alexandrov et al., 2013; Davis et al., 2014; Guo et al., 2013; Huang et al., 2013; TCGA, 2013). Furthermore, large collaborative projects being undertaken are revolutionising our understanding of cancer and other diseases, projects such as TCGA (which permits public access to provide a wealth of WGS data from 34 primary tumour

251

type in over 11,000 cases (Tcga-data.nci.nih.gov, 2015) and the Genomics England 100,000 genome project (Siva, 2015), aim to obtain a comprehensive understanding of the genetics and molecular biology of rare diseases and cancers. As more cancer samples are sequences and NGS continues to improve, so too does our knowledge and understanding of diseases. It is hoped that this added understanding will manifest in improving predictive, diagnostic, prognostic and therapeutic management of cancer patients (Jafri et al., 2015; Rabbani et al., 2014b; Wheeler and Wang, 2013).

5.1.1.4

5.1.2 TERT promoter

While undertaking this study, a high recurrence of somatic telomerase reverse transcriptase (*TERT)* promoter mutations have been reported, in a multitude of cancers, including but not limited to; central nervous system, bladder and thyroid cancers (Vinagre et al., 2013) skin melanoma (Horn et al., 2013; Huang et al., 2013), hepatocellular carcinoma (Nault et al., 2013; Totoki et al., 2014) and urothelial cancer (Borah et al., 2015). *TERT* is involved in maintaining genomic stability and DNA repair; it had been reported to be upregulated in over 90% of cancers with most research focusing on somatic mutations within the *TERT* promoter (Rathmell et al., 2015).

Two mutations in particular, have been highlighted as single nucleotide variants (SNV) hotspots, within the *TERT* promoter 1) C228T the more frequently mutated of the two variants (chr5; 1,295,228 located -124bp upstream of the ATP start site) and 2) C250T (Chr5: 1,295,250 located -146bp upstream of the ATP start site). The SNVs generate new

binding sites for the *TERT* transcription factor motif E-twenty six/ ternary complex factors (Ets/TCF) resulting in an increase of *TERT* promoter activity and in turn increased expression of *TERT*. Which has been associated with poor prognosis, and more advanced forms of cancer (Horn et al., 2013; Huang et al., 2013; Koelsche et al., 2014).

Investigation into the two aforementioned *TERT* promoter SNVs in renal carcinomas, concurs that *TERT* mutations are associated with more aggressive tumours and poor outcome; with Hosen et al., 2015 reporting *TERT* promoter mutations in 6.4% (12/188) ccRCC tested. Wang et al., 2014 identified *TERT* promoter mutations is 9.3% (9/96) ccRCC and 13% (1/8) chRCC they further report that of the 10 mutated tumours, 9 possess the C228T variant, and 1 the C250T variant. A larger cohort of chRCC has since been investigated via the TCGA collaboration. Within this study 4.5% (3/66) were identified with the C228T SNV although the *TERT* mRNA levels for these samples was not increased. Interestingly however via whole genome sequencing, a structural breakpoint was identified 10kbp upstream of the *TERT* transcription start site, through which a further 12% (6/ 50) chRCC tumours were identified to possess a structural variation (e.g. Tandem duplication or translocation) that affected the *TERT* promoter. These samples where also reported to have increased mRNA expression of *TERT* (Davis et al., 2014; Rathmell et al., 2015). However, as yet no study of the *TERT* promoter in renal oncocytoma samples had been conducted. We therefore included a screen of the *TERT* promoter hotspots in our study.

5.1.3 Study's Aim

The key aims of this thesis were to identify and profile the genetic and epigenetic variations that exist between two, histologically similar, sporadic forms of RCC; chRCC and renal oncocytoma. The successful identification of specific epigenetic and genetic differences between these two histologies may then potentially, be used to aid and improve the diagnoses of the specific subtypes, thereby enabling the application of more effective personalised and tailored treatments (section 1.7).

Prior to the beginning of this project, only limited investigations had been performed into the genetics of chRCC and renal oncocytoma. Of the studies that had been published, the main observations were of cytogenetic differences between the two histologies; for example, chRCC is frequently found to harbour loss of chromosomes 1, 2, 6, 10 and 21, while oncocytomas are often associated with loss of chromosome 1 and Y (Faisal et al., 1997; Yusenko et al., 2009) (outlined in section 1.4.2.1.3 and section 1.4.2.1.4). Despite these observations, the chromosomal differences alone were still not sufficient to distinguish clearly between the two subtypes. It was thought that perhaps the identification of tumour-specific variants between the two subtypes might provide a more informative and effective method to differentiate between them (Ng et al., 2014). Consistently, in 2014, the TCGA performed a study, which aimed to identify somatic mutations in chRCC samples. To achieve this, they performed a multiplatform investigation (including data obtained from WGS) on 66 sporadic chRCC samples. Through their study, they identified frequent mutations in the TP53 and PTEN genes, in

addition to frequent alterations to the cancer-associated TERT promoter (outlined in section 5.1.2) (Davis et al., 2014).

In light of the above multi-platform study, a WES investigation of chRCC was not necessary; however, as no mutational analysis had yet been conducted on renal oncocytoma, this led us to question whether we could identify any specific genes that may be playing a role in somatic oncocytomas. We therefore decided to perform whole exome sequencing and bioinformatic analysis on DNA obtained from oncocytoma tumours in an attempt to investigate, identify and characterise specific somatic mutations. Following candidate gene selection, the project then aimed to carry out mutational screening using Sanger sequencing in a larger cohort of patients, in order to identify any potentially recurrent somatic mutations. It was hoped that the successful identification of disease-associated genes and specific pathogenic variants may provide invaluable information regarding the biology, and pathogenesis of these tumours, which may ultimately be of clinical benefit in the future.

5.2 **Results**

5.2.1 **Sample data**

A total of five sporadic oncocytoma samples of Japanese origin; three tumour normal pairs (T1-3 and N1-3) and two tumours only (T4-5), were exome sequenced courtesy of Dr Michael Simpson, Kings College London. However this was achieved in two stages; samples T1, N1, T4 and T5 were exome sequenced first, following some analysis detailed below (round 1), the decision was made to sequence a further two T/N pairs, T2, N2 and T3 and N3. Analysis was conducted on the previously mentioned 'coding' files detailing

the high quality and coding region variants, with additional investigation being conducted upon the raw .BAM and BAM.BAI files via IGV. Firstly the gender of the samples was confirmed, by analysing the number of heterozygous X chromosome variants, in males this is typically <100 and in females >200. N1 and T1, T4 and T5 were correlated with the clinical data and confirmed as female with >200 het X variants (227, 218, 218 and 219 variants respectively). Samples N2 and T2, and N3 and T3 were confirmed as male with <100 het X variants, which corresponds with the known clinical data (53, 44, 63 and 66 variants respectively). Due to the high homology within the Y chromosome to other chromosome both genders identified variants in the Y chromosome (Guo et al., 2014).
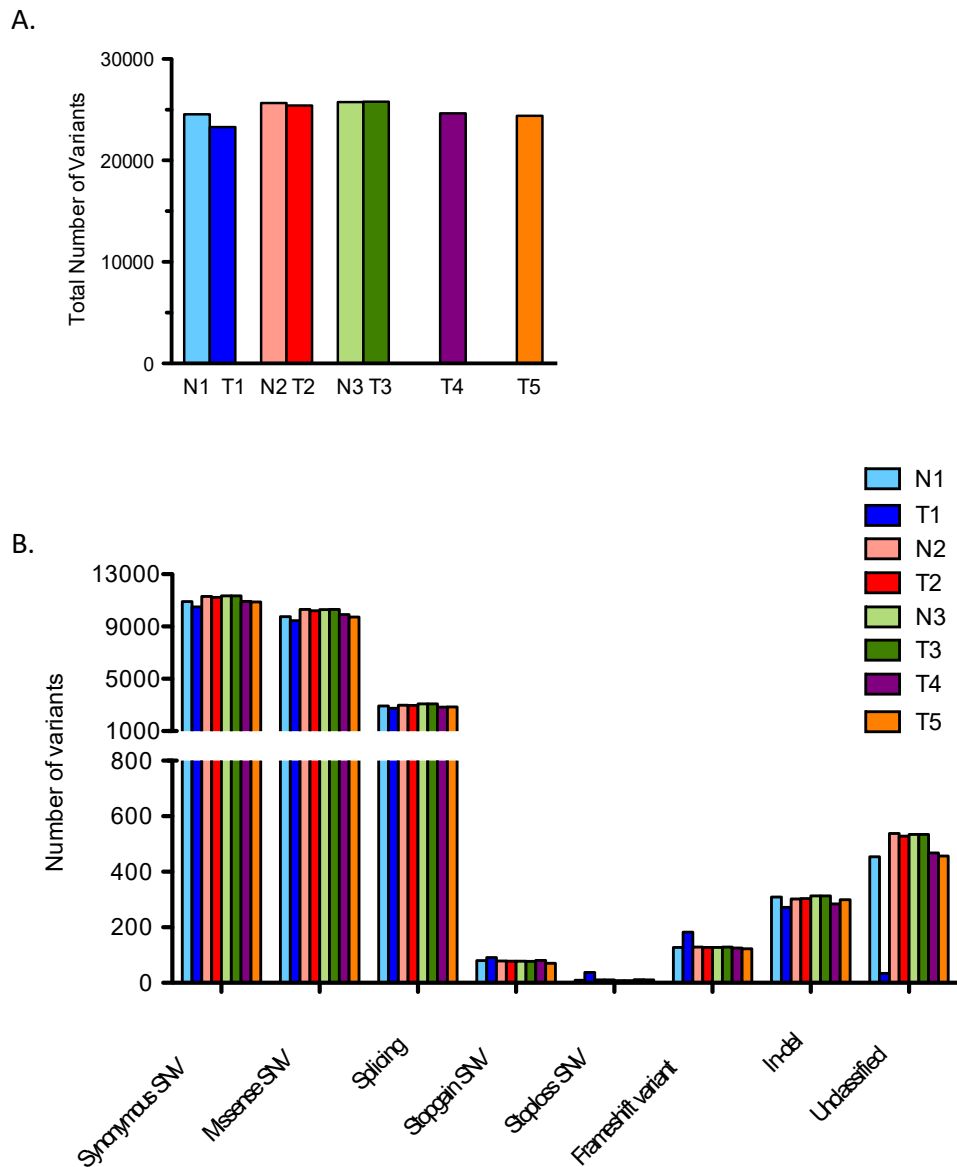
Figure 5-4: Variant Profile of Oncocytoma
Summary of all variants called in the WES coding file (in coding region and high quality).
A) Tally of the total number of variants identified for each sample.
B) Breakdown of the subtypes of variants identified for each samples. Variants lest likely to be deleterious (Synonymous SNV, Missense SNV, and Splicing (intronic variant within 10bp of exon) are shown to be the most common variants identified by WES for all samples.

5.2.1 Variant Profile of Oncocytoma

The total number of variants within the coding regions for each sample is displayed in Figure 5-4 a. A total 199,498 variants were identified, N1= 24,551 variants, T1=23,289 variants, N2=25,645 variants, T2=25,419 variants, N3=25,765 variants, T3=25,790 variants, T4=24,627 variants, T5=24,412 variants. It was noted that slightly more variants were called in the normal samples than the tumour samples; this variation was attributed to a difference in DNA quality, with normal DNA usually being of a higher quality compared to the tumour. As a result during sequencing more variants were likely to pass the phred quality threshold of >20 and therefore be called in the coding file.

Investigation into the types of variants called; show synonymous and missense mutations are the most prevalent in both tumour and normal samples accounting for 84.4% of all variants (44.3% synonymous and 40.07% missense). Splice site variants were called as variants located in the intron within 10bp of the exon boundary, hence the high frequency of 11.7% of total variants; probability dictates most of these variants will have no effect on the protein expression. Variants more likely to be deleterious such as frameshift and nonsense mutations are much less frequent accounting for, 0.5% and 0.3% of the total variants identified (Figure 5-4b).

The high prevalence of synonymous and missense mutations is not unexpected; the majority will have very little or no detrimental effect on the proteins they encode thus having little effect on the well being of the cell. As a result these mutations are likely to be maintained and replicated as the cell undergoes mitosis.

5.2.1 Manual data processing and candidate gene identification

5.2.1.1 *Round 1-Targeted approach*

The original investigation was conducted on the three oncocytoma samples (T1, T4 and T5) and one paired normal (N1), which were in the first batch of exome sequencing.

A manual approach to filter the 72,328 identified variants listed for the three tumours was undertaken as follows and described in Figure 5-5 Firstly any variant present in both the tumours and the normal was removed, a total of 61,065 variants were removed; leaving 11,263 variants in the tumours (T1=539 variants, T4=8,282 variants, T5=2,442 variants). All non-deleterious, synonymous mutations were removed, totalling 1,650 variants removed in this step; leaving 9,613 variants in the tumours (T1=378 variants, T4=4,694 variants, T5=4,541 variants).

Any variants characterised as a SNP with a minor allele frequency (MAF) >1% within dbSNP and ensemble was removed, a total of 7,089 variants were removed leaving 2,524 variants in the tumours (T1=247 variants, T4=1,140 variants, T5=1,137 variants). A MAF of <1% is deemed to be a rare variant (Ramsey et al 2012; Kelly A et al). Thus this step removed any variants common to the general population, as they are unlikely to be pathogenic regarding oncocytoma. To remove any high frequency unclassified SNPs or polymorphisms, or common sequencing errors any variant previously identified more than 10 times 'in-house' by the alignment and variant calling pipeline utilised by Dr Simpson were also removed, totalling 1,071 variants removed; leaving 1,453 variants (T1=53 variants, T4=704 variants, T5=686 variants).
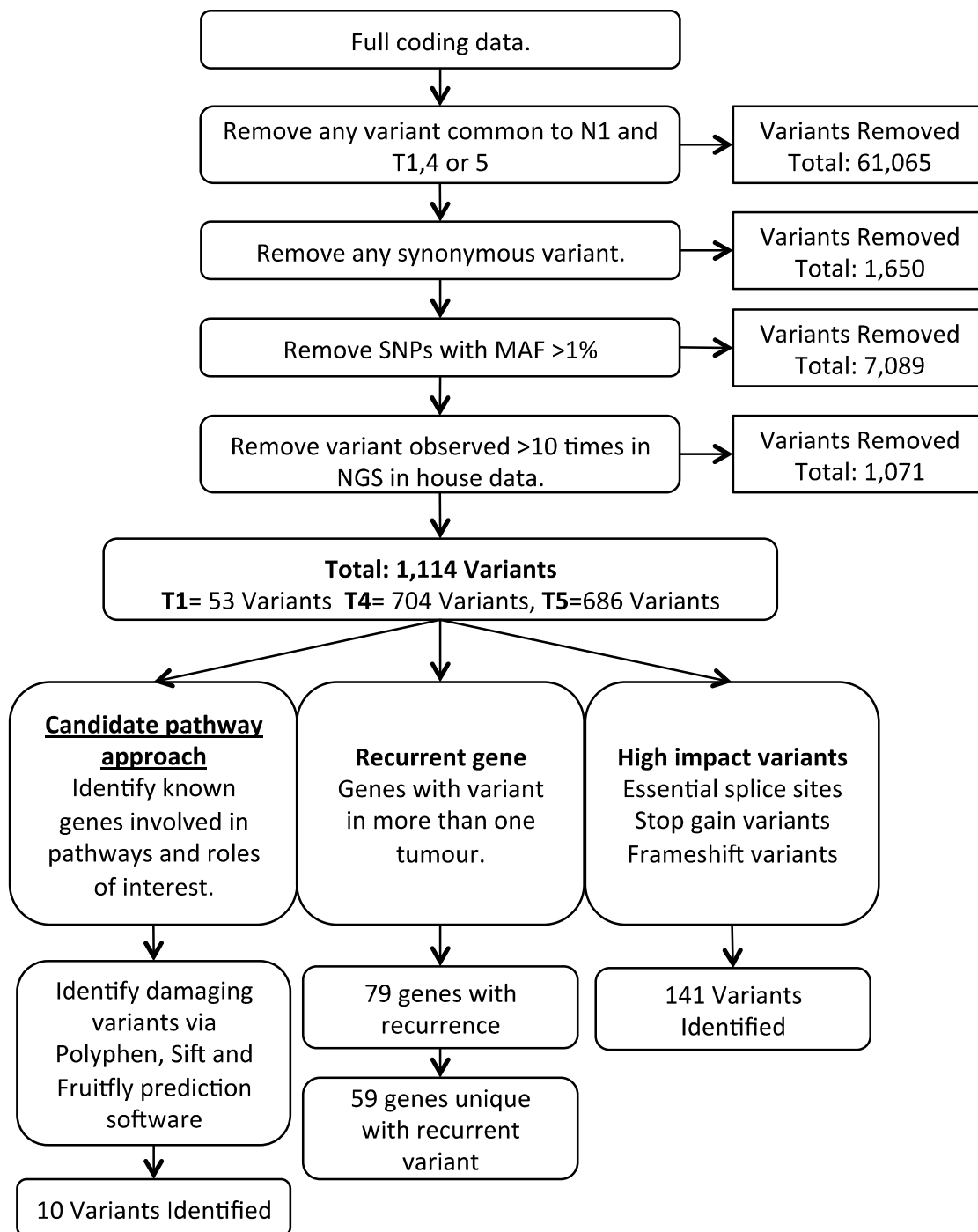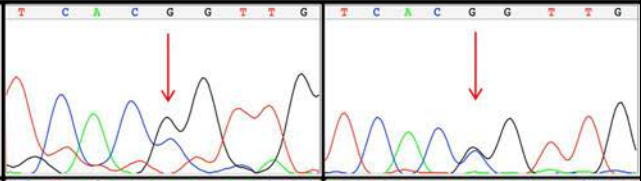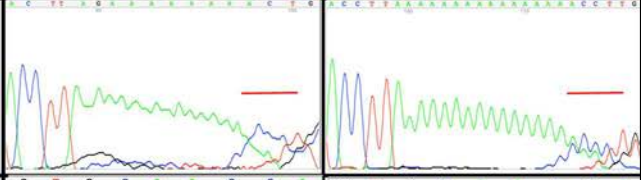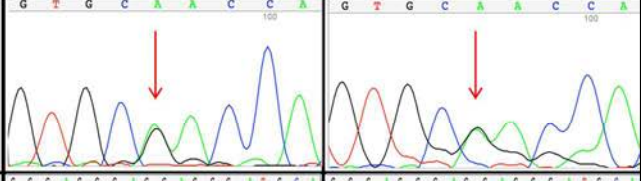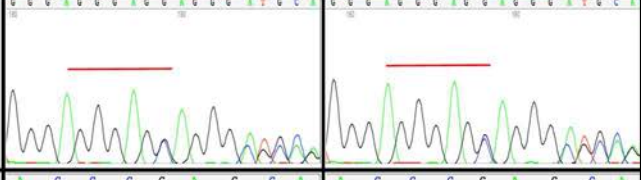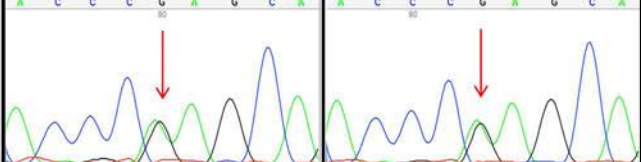
Figure 5-5: Round 1-Targeted approach pipeline
Flow diagram depicting the targeted approach undertaken to identify somatic variants in oncocytoma. Pipeline was applied to WES data from N1, T1, T4 and T5.

To further reduce the gene list, gene prioritisation was initially focused on identifying deleterious variations within key pathways identified by KEGG and Panther with focus directed upon the: AMPK pathway, MAPK pathway, PI3K-AKT pathway, Hippo pathway, TGF-β pathway, p53 pathway, SWI/SNF complex (widely associated with RCC (TCGA, 2013; Varela et al., 2011)) and as well as genes associated with apoptosis and the 3p chromosomal arm, which has previously been associated with cancers and RCCs (Gnarra et al., 1994; Lovell et al., 1999; TCGA, 2013; Zabarovsky et al., 2002). All frameshift and nonsense variants were classified as deleterious, due to the disruption caused on the protein sequence, causing premature termination of the protein, though currently no computational software to predict the severity of these variants is available. For splice site variants, *in-silico* investigation identified the probability of a splice site variant resulting in a loss of splice site was undertaken via prediction software from Berkeley Drosophila Genome Project (Reese et al., 1997). For the missense single nucleotide variants (SNV) further *in-silico* analysis was undertaken using two online prediction software, Sorting Tolerant From Intolerant algorithm (SIFT) (Kumar et al., 2009) and Polyphen (Adzhubei et al., 2010). Both programmes assess and score the probability of an amino acid change impacting on the protein function by assessing protein sequence homology, conservation between species and the physical properties of amino acids. This method yielded 10 variants of interest in 9 genes associated with candidate pathways. *Lamb2* displaying two variants at exon 27 and 24 for sample T4. Via Sanger sequencing, of both the tumour and normal samples, these 10 variants were investigated in-house, all 10 variants were identified to be germline Table 5-1.

| Gene | ID | Exome Sequencing variant | Novel | Pathway/ role | Deleterious | | Sanger Seq. confirm | Tumour | Matched Normal |
					Poly-phen	SIFT			
ADAMTS9	T4	Missense: NM_182920 c.G4240C: p.G1414R	Novel	Chr3p	Y	0.969	0	G	
APAF1	T5	Splicing: c.-3 del 3 AA deletion	In-house: 4	p53	Y	N/A	N/A	G	
BCL11B	T4	Missense: NM_138576 c.G1381A: p.D461N	Novel	SWI/SNF	P	0.582	0.45	G	
CAPN1	T5	Splicing: c. -13 del7 TCCCTCC	Novel	Apoptosis	Y	N/A	N/A	G	
FLNB	T4	Missense: NM_001457 c.G7232A: p.R2411Q	Novel	MAPK and Chr3p	Y	0.522	0.04	G	
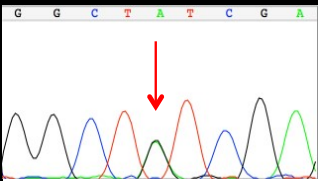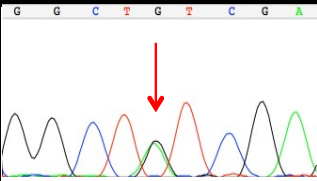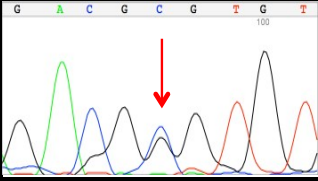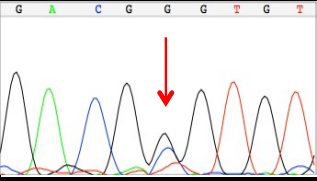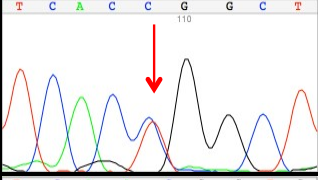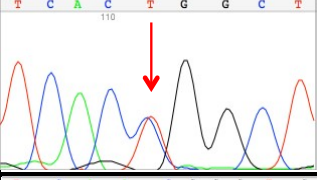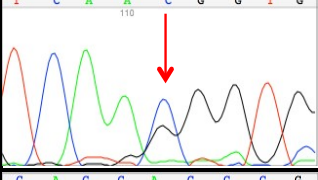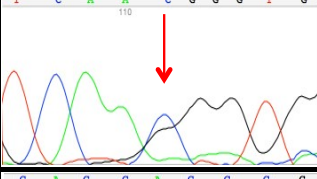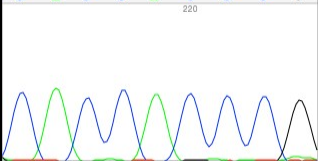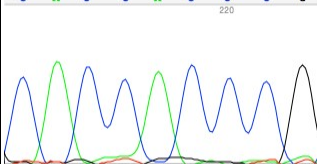
Gene	ID	Exome Sequencing variant	Novel	Pathway/ role	Deleterious		Sanger Seq. confirm	Tumour	Matched Normal
					Poly-phen	SIFT			
LAMB2	T4	Missense: NM_002292 c.G4283A: p.C1428Y	Novel	Chr3p	Y	1	0	G		
LAMB2	T4	Missense: NM_002292 c.G3664C: p.G1222R	Novel	Chr3p	Y	1	0.06	G		
RASGRF2	T4	Missense: NM_006909 c.C2633T: p.P878L	Novel	MAPK	Y	0.998	0.03	G		
SETD2	T4	Missense: NM_014159 c.G407C: p.R136T	Novel	Chr3p - epigenetic regulator of Histone 3 (TCGA, 2013)	Y	0.651	0	G		
SMAD4	T1	Missense: NM_005359 c.A872C: p.H291P	In-house: 8	Hippo and TGFb	Y	0.999	N/A	F		

Table 5-1: Sanger sequencing to confirm somatic status of WES variants associated with known cancer pathways
Table detailing gene name accession number, information on the variant information including type of variant and variant position and base or amino acid change in relation to cDNA (c.) and protein (p.), associated pathway, and the probability of mutation being damaging (Y=damaging, P=Possibly damaging) predicted by SIFT and Polyphen2. Somatic status was confirmed by Sanger sequencing, S= somatic, G=Germline and F=False positive variant called in WES. Variant is highlighted by a red arrow for SNV, or red bar for multiple nucleotide variants.

Focus was then turned to identifying recurrent genes, characterised as genes with multiple variants in different tumour samples. This identified 79 genes presenting with variants in two or more of the oncocytoma samples, of which 21 variants were the same between the samples leaving 59 genes with unique recurrent variants. No gene was identified to possess variants in all three samples and not in the normal.

Another approach undertaken was to look at the high impact variants, here high impact variants were defined as nonsense, splice site mutations <5bp from splice site (investigation into severity via fruitfly.org identified a variant >5bp was unlikely to be damaging) and frameshift variants. From this analysis, a total of 141 variants were identified across the three tumours, (20 nonsense, 20 frameshift and 101 splice sites of which 41 were uncharacterised)

The gene lists returned from these approaches proved to remain immense and unsatisfactory and could not be logically reduced. Additionally previous experience has led us to believe that the majority of variants are likely to be germline, especially when lacking the normal exome sequencing data for two of the samples. The methods undertaken to filter the data possess a bias towards previously characterised genes and pathways and high impact variants, excluding many novel and missense variants. It was therefore decided a further two tumour samples together with corresponding normals would be exome sequenced, to identify novel somatic variants.

5.2.1.2 *Round 2- Manual approach*

The following, manual approach was applied with the aim of identifying novel candidate genes, associated with oncocytoma development while remaining unbiased as to the function and included the exome sequencing data of an additional 2 T/N pairs (T2/N2 and T3/N3) (Figure 5-6)

In a similar starting methodology as was originally undertaken when investigating the candidate genes in round 1, initial filtering omitted any variant common to both tumour and normal samples in all oncocytoma samples sequenced (a total of 115,393 variants were removed in this step; leaving 8,144 variants T1=311 variants, T2=222 variants, T3=290 variants, T4=3,739 variants, T5=3,582 variants). Followed by the removal of any synonymous variant (a total of 3,247 variants were removed; leaving 4,897 variants T1=238 variants, T2=165 variants, T3=219 variants, T4=2,195 variants, T5=2,080 variants).

The next stage was to disregard any variants previously characterised as an SNP with a MAF >1% within dbSNP and ensemble. In addition, all variants recorded >10 times 'in-house' the alignment and variant pipeline were removed. A total of 3,436 probes were removed due to the SNP MAF and in-house observations, 2,795 and 641 variants respectively; leaving 1,461 variants T1=52 variants, T2=45 variants, T3=64 variants, T4=660 variants, T5=640 variants.

It was at this point the filtering method diverged from the processes undertaken in round 1, and the splice site mutations were investigated. As previously mentioned splice sites were investigated via computational prediction analysis program fruitfly.org.

```
┌─────────────────────────────────────┐
│          Full coding data.          │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐       ┌─────────────────────┐
│ Remove any variant present in any│─────▶│  Variants Removed    │
│            N and T               │       │  Total: 115,393      │
└─────────────────────────────────┘       └─────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐       ┌─────────────────────┐
│   Remove any synonymous variant  │─────▶│  Variants Removed    │
│                                  │       │  Total: 3,247        │
└─────────────────────────────────┘       └─────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐       ┌─────────────────────┐
│     Remove SNPs with MAF >1%     │─────▶│  Variants Removed    │
│                                  │       │  Total: 2,795        │
└─────────────────────────────────┘       └─────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐       ┌─────────────────────┐
│  Remove variant observed >10 times│────▶│  Variants Removed    │
│      in NGS in house data.       │       │  Total: 641          │
└─────────────────────────────────┘       └─────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐       ┌─────────────────────┐
│   Remove splice site variants >5bp│────▶│  Variants Removed    │
│  from TSS and any splice site variant│   │  Total: 240          │
│    <5bp predicted to be benign.  │       └─────────────────────┘
└─────────────────────────────────┘
                  │
        ┌─────────┴─────────┐
        ▼                   ▼
┌──────────────────┐  ┌──────────────────┐
│ Matched T/N pairs │◀▶│  Unmatched T     │
│ Total= 107 Variants│  │ Total: 1,114 Variants│
│ T1= 41 Variants   │  │ T4= 571 Variants │
│ T2= 29 Variants   │  │ T5= 543 Variants │
│ T3= 37 Variants   │  └──────────────────┘
└──────────────────┘
        │
        ▼
┌──────────────────────────────┐
│  Assess raw data in IGV and  │
│       compare normal         │
│ Criteria: Variant in <5% N reads│
│   but >15% Tumour Read       │
└──────────────────────────────┘
        │
        ▼
┌──────────────────────────────┐
│          Summary             │
│   28 Variants Somatic        │
│   64 Variants Germline       │
│ 15 Variants did not meet criteria│
└──────────────────────────────┘
```

Figure 5-6: Round 2-Manual approach pipeline
Flow diagram depicting the manual approach undertaken to identify somatic variants in oncocytoma. Pipeline was applied to WES data from N1, T1, N2, T2, N3, T3, T4 and T5.
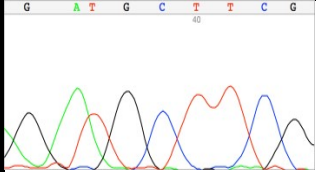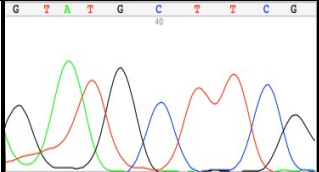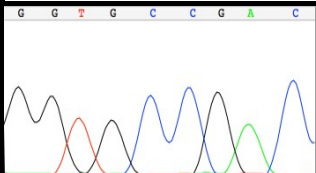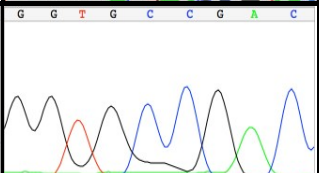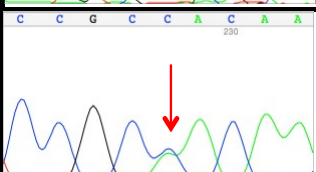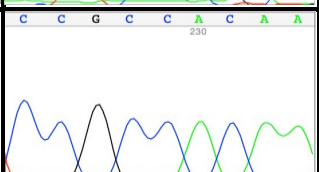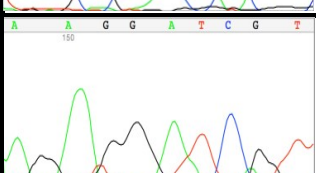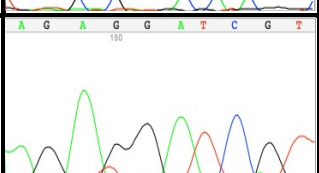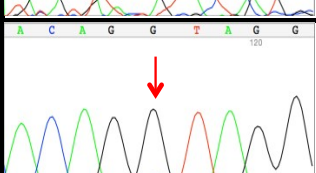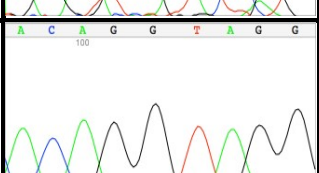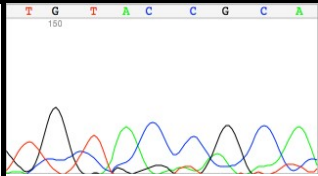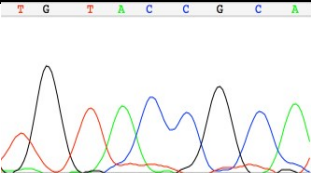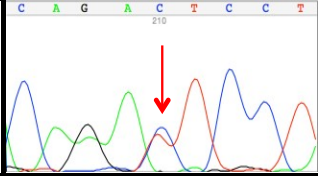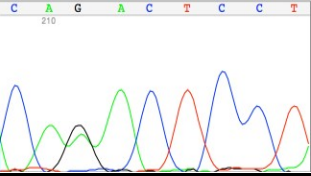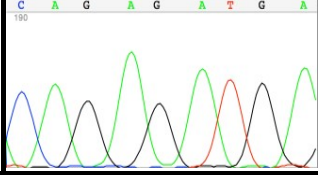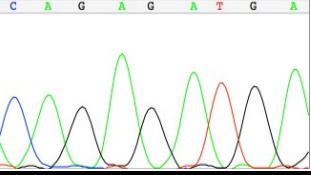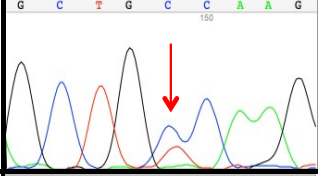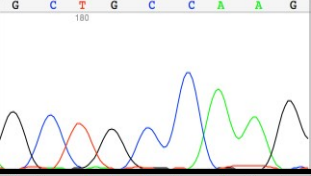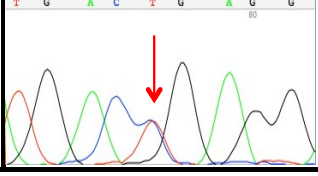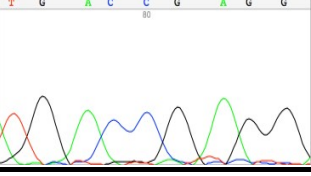
From our previous investigation it was identified as highly improbable that variants >5bp away from the slice site would be damaging, thus all variants >5bp from the splice site were removed. All splice site variants under 5bp were investigated *in-silico* using fruitfly.org, any variant that was predicted to not cause a loss of splice site were removed (a total of 240 variants were removed).
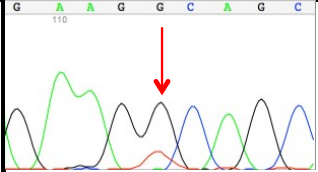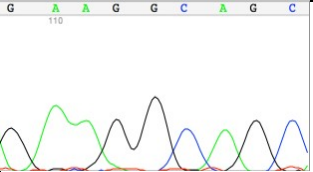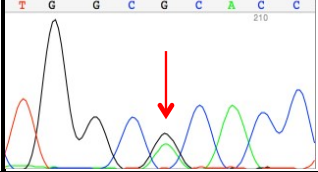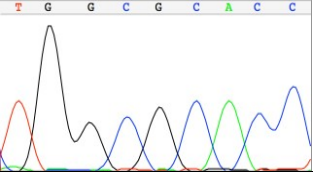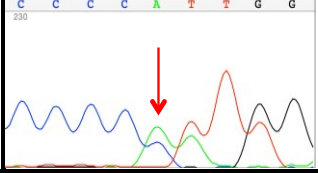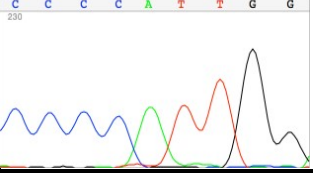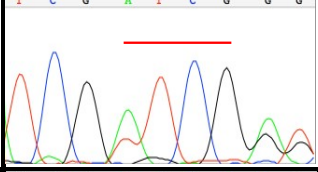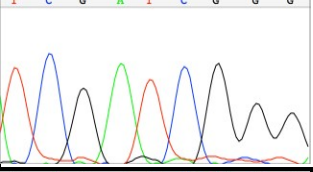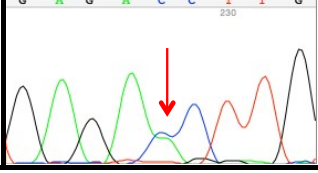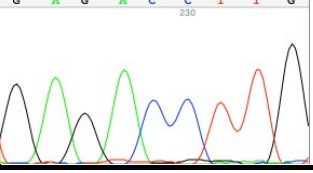
This pipeline identified 1,221 variants of interest distributed accordingly, T1= 41 variants, T2= 29 variants, T3= 37 variants, T4= 571 variants, T5= 543 variants). Due to the sheer volume of variants still remaining in the unmatched tumours, and no logical method to further reduce this number, the following filtration was conducted on all variants identified from T1-3 the tumour normal pairs. For matched tumour normal samples, the raw BAM files for the remaining 107 variants was analysed using IGV (Integrative Genomics Viewer v2.3.40). Read depth and mutation frequency was compared between tumour and normal samples, where variants that were present in >15% tumour reads and <5% Normal reads were deemed to be likely somatic variants and unlikely to be false positives, this is the same criteria applied by Dr Michael Simpson when using the computational pipeline Varscan2 (Koboldt et al., 2012). The result of this process identified 28 somatic variants, 64 germline (Normal >5% of reads), and 15 variants did not meet the criteria. These variants identified as somatic were confirmed via Sanger sequencing on the original samples, of the 28 identified variants; 18 were confirmed as somatic, 4 as germline and 6 false positives (Table 5-2).

Gene	ID	Exome Sequencing variant	Novel	Mut. Reads in IGV files				Sanger Seq.	Tumour	Matched Normal
				T	N1	N2	N3			
AKAP1	T2	Missense: NM_003488 c.G550A: p.V184I	Novel	22.4% (35/156)	0% (0/152)	0% (0/155)	0% (0/178)	S		
CNBD2 C20orf152	T1	Frameshift: NM_080834 c.1322delT: p.L441fs	Novel	44% (11/25)	0% (0/40)	0% (0/104)	0% (0/135)	S		
CACNA1A	T1	Missense: NM_001127222 c.6658_6659insCAC: p.P2220delinsHP	Novel	21.4% (3/14)	1.9% (1/54)	0% (0/43)	0% (0/62)	G		
CRIM1	T2	Missense: NM_016441 c.G2140T: p.V714L	Novel	24.5% (12/49)	0% (0/56)	0% (0/55)	0% (0/83)	S		
CYP51A1	T1	Missense: NM_00078 c.G853T: p.D285Y	rs146408738 (MAF: 0.04%)	50% (3/6)	N/A	0% (0/16)	0% (0/7)	G		

Gene	ID	Exome Sequencing variant	Novel	Mut. Reads in IGV files				Sanger Seq.	Tumour	Matched Normal
				T	N1	N2	N3			
DCAF4L2	T1	Missense: NM_152418 c.A1163G: p.D388G	Novel	46% (29/49)	0% (0/57)	0% (0/67)	0% (0/83)	S		
DERL2	T1	DERL2: NM_016041 c.210_211del: p.70_71del	Novel	42.2% (19/45)	0% (0/61)	0% (0/54)	0% (0/49)	S		
DNAH8	T1	Missense: NM_001206927 c.A2549G: p.E850G	rs117061525 (MAF: 0.91%)	80% (5/4)	0% (0/1)	0% (0/44)	0% (0/76)	G		
ESYT2	T1	Missense: NM_020728 c.A2455G: p.T819A	Novel	27.8% (30/108)	0% (0/64)	0% (0/166)	0% (0/156)	F		
ETFB	T3	Missense: NM_001985 c.T53G: p.V18G	Novel	50.9% (28/55)	0% (0/459)	0% (0/46)	0% (0/61)	S		

Gene	ID	Exome Sequencing variant	Novel	Mut. Reads in IGV files				Sanger Seq.	Tumour	Matched Normal
				T	N1	N2	N3			
FAM157B	T3	Missense: NM_001145249 c.C409T: p.L137F	Novel	50% (2/4)	N/A	N/A	N/A	F		
FMNL1	T1	Missense: NM_005892 c.C1564T: p.P522S	Novel	40% (2/5)	0% (0/13)	0% (0/22)	0% (0/21)	F		
FOXI1	T3	Missense: NM_144769 c.C517A: p.H173N	Novel	47.5% (103/217)	0% (0/60)	0% (0/181)	0.5% (1/199)	S		
GYPB	T3	Missense:NM_002 100 c.T233A: p.I78N	Novel	33.8% (45/133)	0% (0/69)	0% (0/101)	0% (0/113)	F		
KSR1	T2	Splicing: NM_014238 c.635+1G>C	Novel	36.2% (21/58)	0% (0/52)	0% (0.32)	0% (0/41)	S		

Gene	ID	Exome Sequencing variant	Novel	Mut. Reads in IGV files				Sanger Seq.	Tumour	Matched Normal
				T	N1	N2	N3			
MARCH4	T2	Nonsense: NM_020814 c.C897G: p.Y299X	Novel	23.5% (12/51)	0% (0/34)	0% (0/67)	0% (0/70)	F		
MKI67	T1	Missense: NM_001145966 c.C2558T: p.T853I	Novel	41.4% (94/227)	0% (0/348)	0% (0/115)	0% (0/128)	S		
MRC1	T3	Missense: NM_002438 c.G1135T: p.D379Y	Novel	21.1% (4/19)	0% (0/2)	0% (0/27)	0% (0/28)	F		
MRPL9	T1	Missense: NM_031420 c.C770T: p.A257V	Novel	38.8% (40/103)	0% (0/94)	0% (0/59)	0% (0/79)	S		
MYOM1	T1	Nonsense: NM_003803 c.C4300T: p.R1434X	Novel	39.2% (56/143)	0% (0/121)	0.9% (1/114)	0% (0/197)	S		

Gene	ID	Exome Sequencing variant	Novel	Mut. Reads in IGV files				Sanger Seq.	Tumour	Matched Normal
				T	N1	N2	N3			
NOLC1	T2	Missense: NM_004741 c.G703T: p.A235S	Novel	30.1% (22/73)	0% (0/88)	0% (0/84)	0% (0/97)	S		
NPBWR2	T1	Missense: NM_005286 c.G473A: p.R158H	In-house: 1	42.9% (6/14)	0% (0/39)	0% (0/88)	0% (0/126)	S		
OBSCN	T1	Frameshift: NM_001098623 c.21278_21279ins C: p.P7093fs	Novel	36.8% (14/38)	0% (0/157)	0% (0/80)	0% (0/76)	S		
RNF146	T3	Frameshift: NM_001242852 c.980_981insTTCG : p.R327fs,RNF146	Novel	36.2% (55/152)	0% (0/147)	0% (0/149)	0% (0/162)	S		
TEKT5	T1	Missense: NM_144674 c.C977A: p.T326N,	Novel	43.9% (54/123)	0.4% (1/278)	0% (0/166)	0% (0/220)	S		
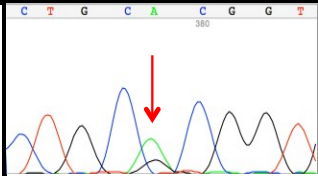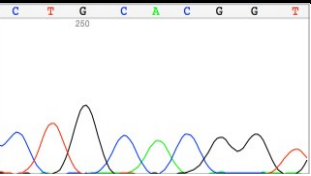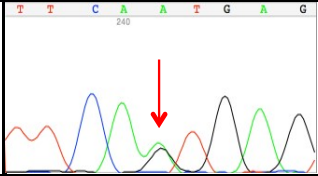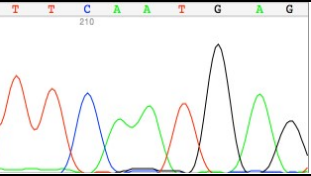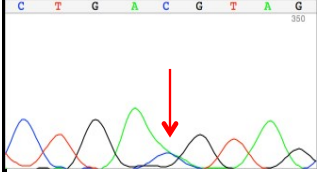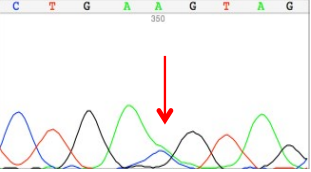
Gene	ID	Exome Sequencing variant	Novel	Mut. Reads in IGV files				Sanger Seq.	Tumour	Matched Normal
				T	N1	N2	N3			
ULK1	T2	Missense: NM_003565 c.A314G: p.H105R,	Novel	22.2% (8/36)	0% (0/76)	0% (0/31)	0% (0/48)	S		
ZC3H7A	T3	Missense: NM_014153 c.A1090G: p.M364V,	rs144891791 (MAF:0.23%)	38.5% (30/78)	0% (0/30)	0% (0/67)	0% (0/68)	S		
ZNF717	T1	Missense: NM_001128223 c.C2694A: p.D898E	In-house: 5	75% (3/4)	N/A	N/A	N/A	G		

Table 5-2: Sanger sequencing to confirm somatic status of WES variants identified through manual pipeline.
Table details gene name and accession number, the sample in which the variant was identified, information on the variant including type of variant position and base or amino acid change in relation to cDNA (c.) and protein (p.), novelty status, summary of IGV files documenting the percentage of reads presenting with the variant. Somatic status was confirmed by Sanger sequencing, where S= somatic, G=Germline and F=False positive variant called in WES. Variant is highlighted by a red arrow for SNV, or red bar for multiple nucleotide variants.

5.2.2 Varscan2

The Varscan2 program is a computation pipeline, which utilises algorithms to determine the genotype of paired normal and tumour sample and call somatic mutations By simultaneously analysing the tumour and normal SAM-tools pileup output (created from the raw BAM files), and performing a direct pairwise comparison of bases called and sequence read depth at each position; Varscan2 is able to identify variations between the normal and tumour sequences (Figure 5-7). If a variation is identified in both the tumour and normal, then it is deemed to be a germline variant. If a variant is identified in one sample and not the other (or low frequency in the other), then the variation is subjected to a one-tailed Fisher's Exact test, where the read counts of the variation (number of reads that present the variation) compared to read counts of wild type (wt) is compared between the tumour and the normal. If the resulting P-value is significant then the variation is deemed to be somatic. If the threshold of significance is not met then the variant is classified germline.

In addition to assessing the somatic status of a variant, Varscan2 can also identify Loss of Heterozygosity (LOH) and Somatic Copy Number Alterations (SCNAs). LOH is calculated by analysis of the variant read counts and comparing between the tumour and the normal. A variant present in >75% of reads was classified as homozygous. LOH was classified as a variant that was homozygous in the tumour and present but heterozygous in the normal. SCNAs are defined as deviations in the log-ratio of the sequence coverage depth within a tumour-normal pair, identified by direct comparison of the normalised sequencing read depth between the tumour and the normal (Koboldt et al., 2012) .
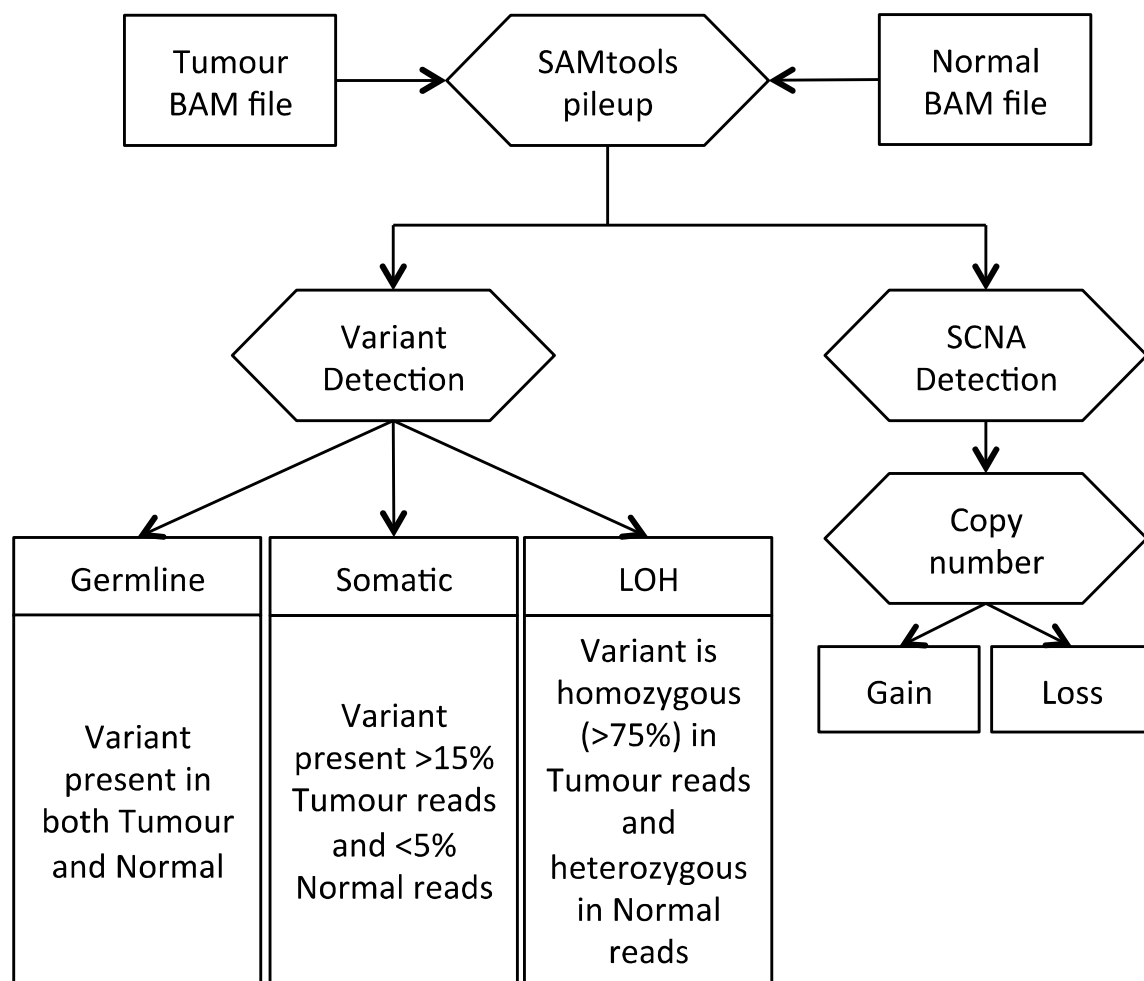
Figure 5-7: Varscan2 pipeline

Flow diagram of the pipeline employed by Varscan2. Adapted from Koboldt et al., 2012.

Further to the aforementioned calculations, Varscan2 also utilises a false-positive filter, which interrogates each variant for nine empirically derived criteria, with the aim of distinguishing true variants from sequencing or alignment related artefacts. In addition to ruling out false positives Varscan2 further categorises somatic variants, into high and low quality; by analysing read depth and mutation frequency and comparing between tumour and normal samples, (similar to the method employed when analysing the IGV files during manual processing). The default filter is set at, variants that were present in >10% tumour reads and <5% Normal reads were deemed to be high quality somatic variants. However, to improve the stringency for the pipeline Dr Michael Simpson used the threshold of >15% tumour and <5%, when analysing our samples.

It should also be mentioned that by default, VarScan2 requires a minimum depth coverage of 3× and a minimum phred base quality of 20, however the higher the coverage and quality of sequencing the less likely it is to yield both false positive and false negative results. Additionally it should also be noted that, due to the calculations undertaken, Varscan 2 is only applicable for tumour normal pairs that have been exposed to identical hybridisation conditions therefore required to have been analysed on the same sequencing array (Koboldt et al., 2012).

As previously mentioned, Dr Michael Simpson kindly ran our data for the tumour/normal pairs (T/N1, 2 and 3) through the Varscan2 pipeline. Adjusting the present criteria to the more stringent >15% tumour reads and <5% normal reads, this identified a total of 318 somatic variants that would be of interest (T1= 293, T2= 12 variants, T3= 13 variants). The discrepancy in somatic variant hits for T1 was

attributed to the fact that N1 and T1 had being sequenced on separate arrays and thus presented a hybridising bias, as a result the varscan2 pipeline identified an unrealistic number of somatic variants for this paired sample. When comparing the manually identified and validated genes to the Varscan2 list, 94% (17 out of 18) of the confirmed somatic genes were also identified by Varscan2 pipeline (Figure 5-8).
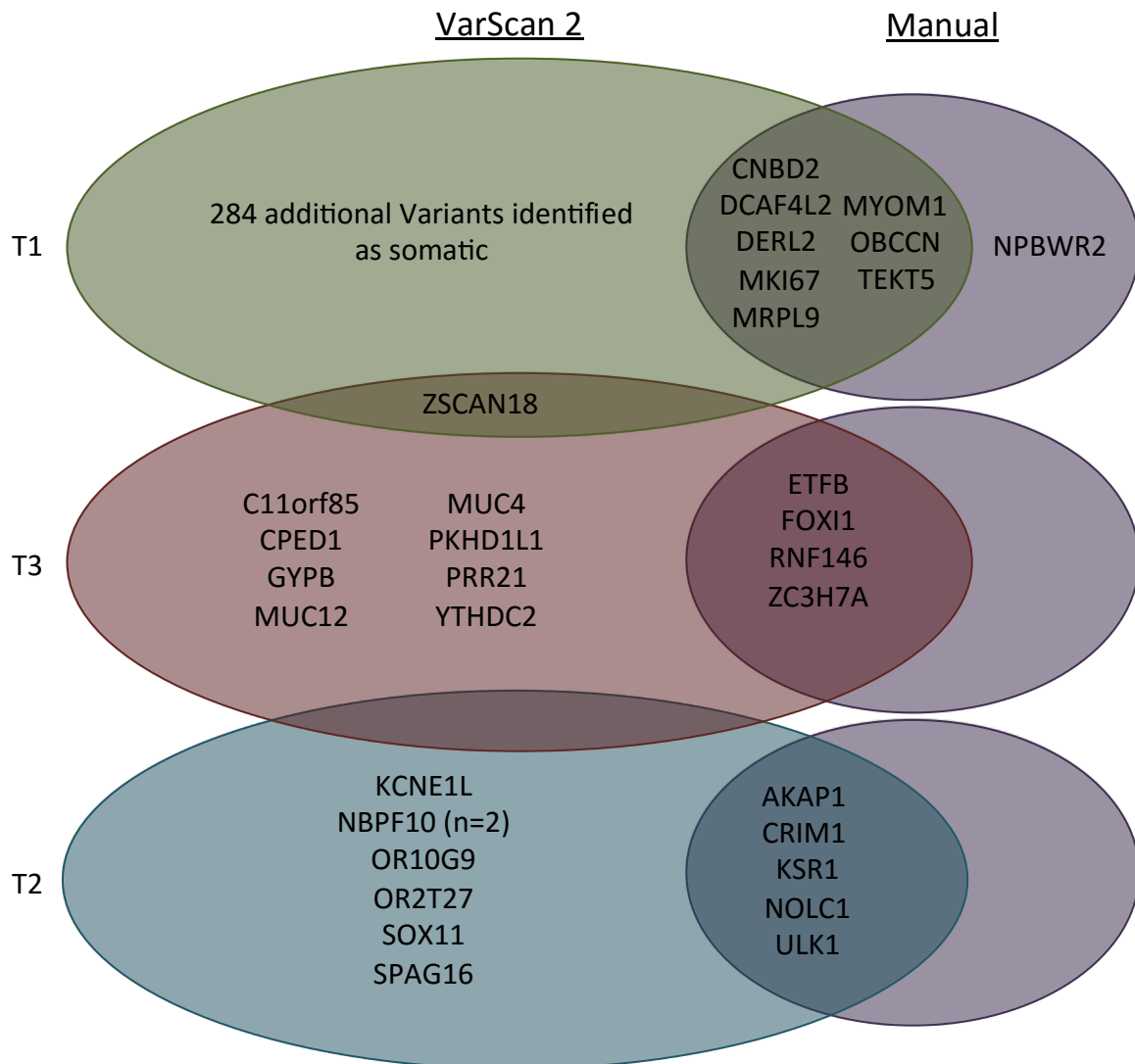
Figure 5-8: Varscan2 gene list compared to the validated genes from the manual pipeline

Venn diagram comparing Varscan2 predicted somatic genes to the validated and confirmed somatic genes identified via the manual pipeline for each T/N pair. 94% (17/18) confirmed somatic genes were also identified by Varscan2.

5.2.3 The candidate genes

Of the 18 confirmed somatic variants, 1 was a stop gain variant (*MYOM1)*, 4 were frameshift variants (*RNF146, DERL2, CNBD2 (C20orf152), OBSCN),* 1 splice site variant (*KSR1)* and 12 were missense SNV (*AKAP1, CRIM1, DCAF4L2, ETFB, FOXI1, MKI67, MRPL9, NOLC1, NPBWR2, TEKT5, ULK1, ZC3H7A).* The predicted effect of variant, function of gene and prioritisation of each candidate gene is summarised in Table 5-3.

Once the candidate gene list was obtained, prioritisation was then performed. Initially the list was screened against the whole exome data of the two unpaired samples, to establish if any genes had recurring variants, in more than one sample. This process identified two additional variants in *NOLC1* and *KSR1.* An indel (non-frameshift) deletion of TCT in T4 for *NOLC1* c.1132_1134del:p.378_378del. For *KSR1* a non-synonymous variant in T5 was identified; *c.C629T:p.T210M.*

Next we identified the probability of a variant being deleterious to the protein. As previously mentioned in the round 1 filtration, the splice site variant was previously predicted to be damaging during the filtering process while all frameshift and nonsense variants were classified as damaging, due to the resulting truncation of the protein sequence.

For the three somatic frameshift variants, it was possible to calculate how many amino acids (aa) were altered from the original sequence before the truncation of protein sequence. The frameshift of *RNF146,* (NM_001242852) c.980_981insTTCG resulted in 4aa change including the stop codon that truncated the protein sequence (FS4). As a result RNF146:FS was truncated 27aa prematurely at p.332 while ordinarily RNF146:wt

terminates at p.359. For *DERL2, (NM_016041) c.210_211del* a deletion of two TT nucleotides results in a frameshift, and FS2 (2 amino acids changed prior to truncation of the sequence) thus DERL2:FS truncated 167aa prematurely at p.73, in the wild type DERL2:wt terminates at p.240. *CNBD2 (C20orf152), (NM_080834) c.1322delT* causing FS5 (5 amino acids changed prior to truncation of the sequence), resulting in CNBD2:FS truncating at p.445, 128aa premature, CNBD2:wt terminated at p.573.

For the missense SNV, the predicted severity was assessed following the aforementioned methods for Polyphen2 and SIFT. Of the 12 missense candidate variants that were investigated 4 variants were predicted to be damaging in both SIFT and Polyphen (*DCAE4L2*, *ETEB, FOXI1, NPBWR2*), 5 variants were predicted to be damaging in one of the two software and were therefore classified in this study as potentially damaging (*CRIM1, MKI67*, *MRPL9*, *NOLC1, ULK1),* while 3 variants (*AKAP1, TEKT5, ZC3H7A*) were predicted to be tolerated in both prediction programs, and were therefore not investigated further.

In addition to identifying the impact of the variant on the protein, each gene was also assessed to ascertain whether mutations have previously been identified in other cancers (via COSMID and cBioportal analysis of TCGA data), the prevalence of high-impact mutations within the gene in a general population (Ensembl; which encompasses, Exome Variant Server and, 1000 genome project, Hapmap project multiple observation and COSMIC, in addition the Japanese SNP database and ExAC and the known function of the gene (Table 5-3).

Gene	Confirmed Somatic WES variant	Recurrent WES variant	Pathway/ Role (Uniport.org)	Deleterious		Database Search	
				Polyphen	SIFT	COSMIC	Ensembl
NOLC1	**T2: MS:** NM_004741 c.G703T: p.A235S	**T4: Indel:** c.1132_1134del: p.378_378del	**Nucleolar and coiled-body phosphoprotein**; Involved in snoRNA formation and regulation of MDM2 transcription (Hwang et al 2009)	P 🟢 0.01	🔴 0.03	**SG**:6(5) **FS**:8(5) **ID**:3(1) **MS**:72(45) **Sy**:17(15)	**SG**:16 **FS**:26 **S**:24 **SR**:95 **InD**:57 **MS**:679 **Sy**:259
KSR1	**T2: S:** NM_014238 c.635+1G>C	**T5: MS:** c.1132_1134del: p.378_378del (SNP (uncharacterised) Frequency 0.000239)	**Kinase Supressor of Ras1:** Scaffolding protein connecting MEK to RAF, promotes activation of the NOTCH signalling pathway in several cancers.	Y N/A	N/A	**SG**:7(0) **FS**:24(6) **MS**:103(73) **Sy**:36(27)	**SG**:37 **FS**:36 **S**:13 **SR**:145 **MS**:975 **Sy**:490
ULK1	**T2: MS:** NM_003565 c.A314G: p.H105R	-	**Unc-51 Like Autophagy Activating Kinase 1:** Regulates autophage in response to starvation, regualted by mTOR and AMPK pathways. Regulates FLCN in BHD	P 🔴 0.98	🟢 0.55	**SG**:6(5) **FS**:12(3) **MS**:119(77) **Sy**:61(44)	**SG**:4 **FS**:7 **S**:4 **InD**:4 **MS**:250 **Sy**:189
CNBD2 C20orf152	**T1: FS:** NM_080834 c.1322delT: p.L441fs	-	**Cyclic nucleotide binding domain containing 2**	Y N/A	N/A	**SG**:5(2) **MS**:71(49) **Sy**:30(21)	**SG**:41 **FS**:19 **S**:15 **MS**:522 **Sy**:148
DCAF4L2	**T1: MS:** NM_152418 c.A1163G: p.D388G	-	**DDB1 and CUL4 associated factor 4-like 2.**	Y 🔴 1.00	🔴 0.03	**SG**:15(6) **MS**:243(139) **Sy**:81(50)	**SG**:20 **FS**:1 **InD**:1 **MS**:217 **Sy**:83
DERL2	**T1: FS:** NM_016041 c.210_211del: p.70_71del	-	**Derlin 2**: Associated with degradation of misfolded glycoproteins and may play a role in cell proliferation	Y N/A	N/A	**SG**:3(2) **FS**:1(0) **MS**:20(10) **Sy**:6(3)	**SG**:12 **SR**:87 **MS**:131 **Sy**:84
ETFB	**T3: MS:** NM_001985 c.T53G: p.V18G	-	**Electron-transfer-flavoprotein, beta polypeptide:** involved in mitrochondrial electron transfer.	Y 🔴 0.85	🔴 0.01	**SG**:2(2) **FS**:5(3) **ID**:1(1) **MS**:30(16) **Sy**:18(10)	**SG**:11 **FS**:13 **S**:6 **SR**:24 **InD**:10 **MS**:244 **Sy**:127

Gene	Confirmed Somatic WES variant	Recurrent WES variant	Pathway/ Role (Uniport.org)	Deleterious			Database Search	
					Polyphen	SIFT	COSMIC	Ensembl
FOXI1	**T3: MS:** NM_144769 c.C517A :p.H173N	-	**Forkhead box I1:** Transcriptional activator involved with kidney devlopment	Y	⬤ 0.99	⬤ 0.00	**SG**:4(4) **FS**:4(1) **MS**:86(54) **Sy**:53(37)	**SG**:3 **FS**:7 **S**:1 **SR**:6 **MS**:184 **Sy**:121
MYOM1	**T1: SG:** NM_00380 c.C4300T: p.R1434X	-	**Myomesin 1**: forms myofibril M line in muscle.	Y	N/A	N/A	**SG**:23(12) **FS**:5(2) **MS**:212(119) **Sy**:78(56)	**SG**:69 **FS**:21 **S**:44 **SR**:198 **InD**:3 **MS**:1164 **Sy**:504
NPBWR2	**T1: MS:** NM_005286 c.G473A: p.R158H	-	**Neuropeptides B receptor:** neuroendocrine system regulator. May be involved with MAPK mediated proliferation in adrenocortical cancers (Andreis et al 2005).	Y	⬤ 0.99	⬤ 0.00	**SG**:2(1) **FS**:1(1) **ID**:1(0) **MS**:59(44) **Sy**:26(20)	**SG**:3 **FS**:1 **InD**:2 **MS**:105 **Sy**:69
OBSCN	**T1: FS:** NM_001098623 c.21278_21279insC: p.P7093fs	-	Obscurin; signaling protein associated with titin, and may have a role in the organization of myofibrils during assembly.	Y	N/A	N/A	**SG**:76(39) **FS**:45(19) **ID**:2(1) **MS**:777(480) **Sy**:402(251)	**SG**:732 **FS**:391 **S**:313 **SR**:866 **InD**:34 **MS**:15128 **Sy**:8151
RNF146	**T3: FS:** NM_001242852 c.980_981insTTCG: p.R327fs	-	**Ring Fingure Protein 146**: E3 ubiquitin-protein ligase,Maybe involved in cell survival and proliferation. Overexpression associated with increased proliferation in lung cancer (Song et al 2014)	Y	N/A	N/A	**SG**:5(3) **FS**:4(4) **MS**:43(29) **Sy**:13(8)	**SG**:20 **FS**:6 **SR**:28 **InD**:5 **MS**:314 **Sy**:130
CRIM1	**T2: MS:** NM_016441 c.G2140T: p.V714L	-	**Cysteine Rich Transmembrane BMP Regulator 1**: May be involved in tissue development and capiliaryformation after TGFβ activation.	P	⬤ 1.00	⬤ 0.28	**SG**:10(5) **FS**:3(3) **MS**:116(76) **Sy**:42(30)	**SG**:11 **FS**:8 **S**:9 **SR**:38 **InD**:3 **MS**:300 **Sy**:166

Gene	Confirmed Somatic WES variant	Recurrent WES variant	Pathway/ Role (Uniport.org)	Deleterious		Database Search	
				Polyphen	SIFT	COSMIC	Ensembl
MKI67	**T1: MS:** NM_001145966 c.C2558T: p.T853I	-	**Marker of Proliferation Ki-67:** Nuclear proteinassociated with cell proliferation.	P 1.00	0.06	**SG**:41(15) **FS**:31(16) **ID**:3(2) **MS**:348(216) **Sy**:137(73)	**SG**:124 **FS**:96 **S**:9 **SR**:53 **InD**:18 **MS**:2567 **Sy**:952
MRPL9	**T1: MS:** NM_031420 c.C770T: p.A257V	-	**Mitochondrial ribosomal protein L9.**	P 0.79	0.79	**FS**:1(1) **MS**:26(13) **Sy**:11(7)	**SG**:4 **FS**:6 **S**:6 **SR**:64 **MS**:109 **Sy**:44
AKAP1	**T2 MS:** NM_003488 c.G550A: p.V184I	-	**A Kinase Anchor protein 1:** binds PKA to the mitrochondria.	T 0.10	0.33	**SG**:8(8) **FS**:6(4) **ID**:2(1) **MS**:111(56) **Sy**:30(18)	**SG**:84 **FS**:39 **S**:23 **SR**:94 **InD**:43 **MS**:1559 **Sy**:900
TEKT5	**T1: MS:** NM_144674 c.C977A: p.T326N	-	**Tektin 5:** Structural component of ciliary and flagellar microtubules.	T 0.00	0.50	**SG**:2(0) **FS**:2(0) **MS**:92(64) **Sy**:40(25)	**SG**:8 **FS**:5 **S**:2 **SR**:36 **MS**:271 **Sy**:126
ZC3H7A	**T1: MS:** NM_014153 c.A1090G: p.M364V	-	**ZincFingure CCCH- Type Containing 7A**: predicted to bing DNA.	T 0.00	1.00	**SG**:6(3) **FS**:3(1) **ID**:1(1) **MS**:83(55) **Sy**:33(21)	**SG**:18 **FS**:8 **S**:20 **SR**:177 **InD**:4 **MS**:657 **Sy**:338

Table 5-3 : Candidate genes identified and confirmed as somatic in oncocytoma

List of the 18 confirmed somatic genes identified via WES of oncocytoma. Table details the gene name, function and accession numbers; type of variant (see below for definitions), variant position and base or amino acid change in relation to cDNA (c.) and protein (p.), samples in which the mutation was identified, the variant identified, the probability of variant being damaging (**Y:** Damaging, predicted to be damaging in both Polyphen and SIFT; **P:** possibly damaging, predicted to be damaging in either Polyphen or SIFT, **T:** Tolerated, mutation predicted as benign in both Polyphen and SIFT). Also listed is the frequency of mutations identified for the gene in COSMIC and Ensemble (Both include TCGA data).

(**MS:** missense; **FS:** Frame shift; **SG:** Stop Gain; **S:** Splicing; **SR**: Splice region (within 10bp of intron) **Sy:** Synonymous)

Notably *KSR1* and *ULK1* showed to have potentially interesting functions regarding a role in cancer.

Firstly, *Kinase Suppressor of RAS1 (KSR1)* has been identified as a scaffold protein, facilitating signal transduction within the Ras signalling pathway leading to ERK/MAPK activation by bringing phosphorylated Raf1, into close proximity to MEK thus allowing the activation of ERK/MAPK complex and transcription (Kolch, 2005).

KSR1 (Kinase Suppressor of RAS 1) has been reported to present oncogenic activity via its involvement in ERF/MAPK activation in Ras-dependent cancers such as pancreatic cancer and lung carcinomas (H. Zhang et al., 2015). Recently reports suggest that high expression of *KSR1* to be associated with upregulation of *NOTCH* and activation of NOTCH signalling in papillary thyroid cancer (Lee et al., 2015). A study has also suggested that *KSR1* inhibition may the method of action and a target for a current anti-cancer drug Terbinafine. The loss of *KSR1*, results in reduced Raf-MEK-ERK signalling and subsequently resulting in an increase in apoptosis (Li et al., 2013).

However for non- Ras-mediated cancers such as breast cancer, it has been reported that KSR1 may act as a tumour suppressor gene, preventing the ubiquination and degradation of TSG *BRCA1*, and correlating with increased survival (Stebbing et al., 2015; H. Zhang et al., 2015), suggesting KSR1 has multiple and complex roles within cancer.

ULK1 (Unc-51 Like autophagy activating Kinase 1) is a serine/threonine kinase involved in driving autophagy and mitophagy of defective cells component and mitochondria respectively, although as yet the underlying mechanisms and substrates have yet to be

confirmed. *ULK1* activation is regulated by several cancer-associated pathways including AMPK and MTOR signalling pathways (Nazio et al., 2013; Petherick et al., 2015; Tian et al., 2015). Additionally *ULK1* has been recently associated with the regulation of the TSG *FLCN* (frequently mutated within BHD renal tumours) by inducing FLCN phosphorylation with Dunlop et al., 2014 reporting a disruption in autophagy within the BHD associated renal tumours investigated.

This analysis, allowed for us to prioritise the screening of the candidate genes;

1. *NOLC1* and *KSR1*: Identified to have mutations in more than one sample.

2. *ULK1:* Selected due to function particularly the association with *FLCN*, and cancer-associated pathways AMPK and MTOR.

3. *CNBD2 (C20orf152), DCAF4L2, DERL2, ETFB, FOXI1, MYOM1, NPBWR2, OBSCN* and *RNF146:* Predicted to be damaging in both SIFT and Polyphen.

4. *CRIM1*, *MKI67, MRPL9*: Predicted to be damaging in one of the two prediction software

5.2.4 Screen of candidate genes

All exons of the candidate genes were screened via Sanger sequencing in 40 additional oncocytoma samples, of either Germanic (n=32) or Japanese (n=8) ethnic origin.

To preserve the supply of patient DNA, samples were whole genome amplified (WGA) prior to the Sanger sequencing screen. Analysis was initially conducted on mutation surveyor with a manual visual assessment using 4-peaks to visually confirm the mutation surveyor report. Inconclusive sequences or sequences with unacceptably high

background noise were re-sequenced in the alternative direction, while any potential variant was first mapped in Ensembl and COSMIC to identify whether variant had been previously reported and whether the variant was non-synonymous. Following, these checks potentially novel variants were re-sequenced in the original unamplified tumour DNA and corresponding normal where possible.

Dr. James Whitworth (University of Birmingham) screened the WGA DNA, for genes *FOXI1, RNF146, ETFB, DERL2* with any potential variant or inconclusive result being repeated by myself on the original stock DNA. Where a variant was confirmed, sequencing of the matched normal (if available) was conducted to test whether the variant was somatic, after first confirming within ensemble that variant was neither a SNP nor synonymous variant.

For *NOLC1* and *KSR1* the status of the recurrent variant identified by WES was investigated in stock tumour DNA and corresponding normal, (Figure 5-9) however both variants were identified as germline. It should also be noted that since receiving the exome data *NOCL1* variant c.1132_1134del:p.378_378del, has since been reclassified from a novel variant to a known SNP rs548784326; with a MAF of <0.01 globally (1000 genome project, and an MAF of 0.02 (4/104 genotypes) in Japanese cohort.

A. **_NOLC1_**: Indel variant identified in WES for T4- **Germline**
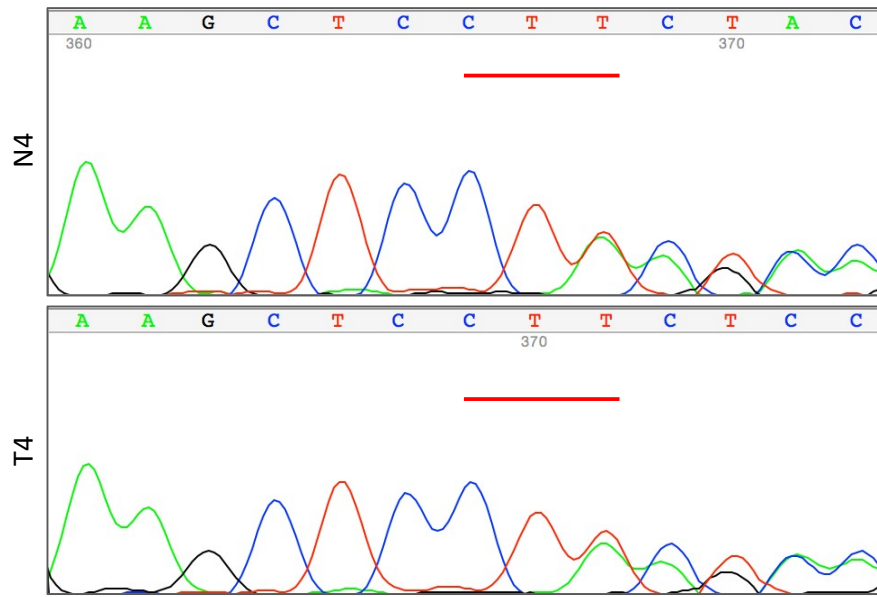
B. **_KSR1_**: Missense variant identified in WES for T5 – **Germline**
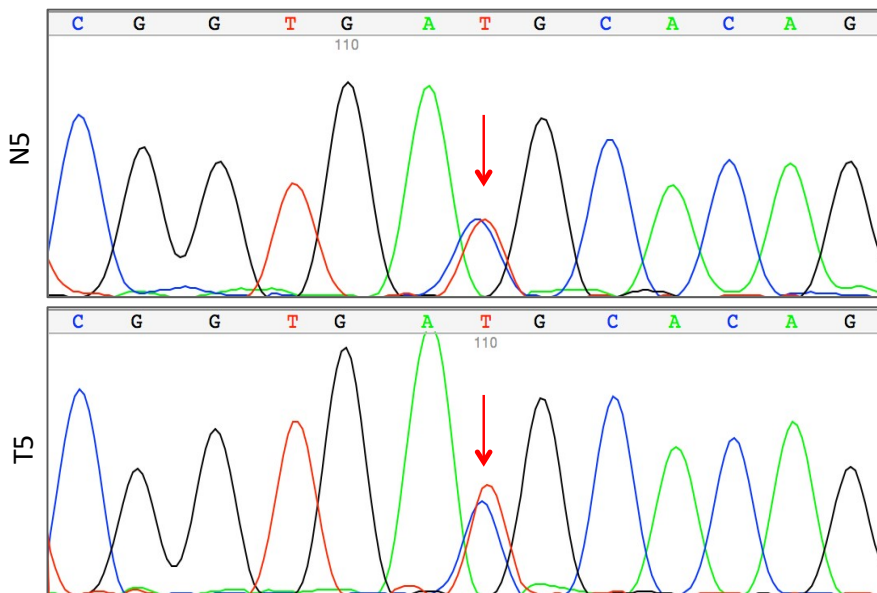
Figure 5-9: Electropherogram of WES recurrent variants _NOLC1_ and _KSR1_
Sanger sequencing showing both variants identified as recurrent in the WES files are germline in the additional sample. **A)** _NOLC1_; variant c.1132_1134del:p.378_378del, reclassified as SNP: rs548784326 (MAF <0.01). **B)** _KSR1_; variant c.1132_1134del: p.378_378del (in WES identified as unclassified SNP with a observational frequency of 0.000239

CNBD2 (C20orf152), DCAF4L2, DERL2, ETFB, FOXI1, KSR1, MYOM1, NOLC1, NPBWR2, RNF146, ULK1 were successfully screened and confirmed in the stock DNA. However, *OBSCN, CRIM1, MKI67* and *MRPL9* were unfortunately unable to be screened due to time constraints within the project.

Results of the genes screened are displayed in Table 5-4. For *CNBD2 (C20orf152), CRIM1, DCAF4L2, DERL2, ETFB, KSR1, MYOM1, NOLC1, NPBWR2, RNF146* and *ULK1,* no further novel damaging mutations were confirmed. Instead only synonymous and characterised SNPs were identified.

Gene	Exons	Additional sample screen			
NOLC1	13	rs116928523 (p.L14L MAF:<0.01(A)) Onco=1/40	rs1049455 (p.S486P MAF:0.18(C)) Onco=8/40		
KSR1	22	rs2293180 (p.A85V MAF:0.17(T)) Onco= 10/40	rs2293181 (p.L96L MAF:0.38(T)) Onco=23/40 (7 Hom)	rs55646895 (p.Q411H MAF:0.01(C)) Onco=1/40	rs55686363 (p.H420H MAF:0.01(T)) Onco=8/40
ULK1	28	rs140040966 (p.G23G) Onco= 3/40	rs142265536 (p.P594P MAF:<0.01(G)) Onco= 1/40	rs77188001 (p.P609P MAF:0.02(T)) Onco= 1/40	rs369546950 (p.N262S) Onco= 1/40
		rs11546872 (p.S1042S MAF:0.03(A)) Onco= 1/40	rs55815560 (p.S665L MAF:<0.01(T)) Onco= 1/40	rs61731334 (p.P662P MAF: 0.01(T)) Onco= 1/40	(c.T1515C p.S505S) Onco= 1/40
		rs511609348 (p.T816A MAF:0.01(A)) Onco= 40/40 (28 Hom)			
CNBD2	12	rs17347958 (p.R37C MAF:0.06(A)) Onco= 7/40	rs138045142 (p.R139H MAF:<0.01(A)) Onco= 1/40	rs6142471 (p.T208A MAF:<0.01(G)) Onco=1/40	rs6060750 (p.P375S MAF:0.24(T)) Onco= 16/40
DCAF4L2	1	No variants identified			
DERL2	7	No variants identified			
ETFB	6	rs533794107 (p.I297I MAF:<0.01(A)) Onco= 10/40 (5Hom)	rs140608276 (p.V79I MAF:<0.01(T)) Onco= 2/40	rs144991327 (p.P52P MAF:<0.01(A\)) Onco= 5/40	rs111454736 (5'UTR) Onco= 2/40

Gene	Exons	Additional sample screen			
FOXI1	2	rs535372917 (p.G40G) Onco= 1/40	rs2277944 (p.R93R MAF:0.31(A)) Onco 13/40 (6 Hom)	**c.C517A (p.H173N)** **Onco= 3/40**	COSM736784 (p.L177L) Onco= 1/40
		c.G505A (p.A154T) Germline Onco= 1/40	rs115399307 (p.T226I MAF:<0.01(T)) Onco= 1/40	rs35678180 (p.S242S MAF:<0.01(T)) Onco= 1/40	rs147596900 (p.A275A MAF:<0.01(T)) Onco= 1/40
		rs56128152 (p.P324P MAF:0.01(C)) Onco= 3/40	rs10063424 (p.Y348Y MAF:0.91(C)) Onco= 39/40 (32 Hom)		
MYOM1	38	rs9964300 (p.A64A MAF:0.07(A)) Onco= 1/40	rs1962519 (p.S181P MAF:0.43(G)) Onco= 19/40	rs201409582 (p.L274L) Onco= 11/40	rs2230162 (p.G333G MAF:0.29(T)) Onco= 18/40 (5 Hom)
		rs8099021 (p.G341A MAF:0.74(G)) Onco= 37/40 (21 Hom)	rs189973743 (p.T387I MAF:<0.01(A)) Onco= 5/40 (1 Hom)	rs11659820 (p.R464R MAF:0.22(T)) Onco= 8/40	rs369493400 (p.G668D MAF:<0.01 (T)) Onco= 1/40
		rs1071600 (p.I960T MAF:0.20(G)) Onco= 15/40	rs3765623 (p.D1408N MAF:0.07(T)) Onco= 1/40	rs187108957 (p.F1558L MAF:<0.01(G)) Onco= 11/40 (1 Hom)	rs2230164 (p.G1680G MAF:0.57(G)) Onco= 30/40
		rs376007739 (p.E1571E) Onco= 12/40	rs2230163 (p.H382H MAF:0.30(A)) Onco= 18/40 (5 Hom)	rs2230167 (p.N446N MAF:0.26(A)) Onco= 16/40 (14 Hom)	
NPBWR2	1	rs150422039 (p.P42L MAF:<0.01(A)) Onco= 1/40	rs41278242 (p.P99P MAF:0.01(A)) Onco= 1/40	rs13036542 (p.P289P MAF:0.47(A)) Onco= 35/40	
RNF146	5	No variants identified			

Table 5-4 Results of targeted sequencing in an addition 40 oncocytoma samples
Results from the targeted sequencing of 40 additional oncocytoma samples. All exons of genes were screen by Sanger Sequencing. Variants identified where mapped to Ensemble for further identification. Most variants were identified to be synonymous and/or a characterised SNP. Within *FOXI1* one variant c.C517A p.H173N was identified to be novel recurrent variant (marked in bold).

5.2.5 FOXI1

FOXI1, was identified to possess a somatic SNV predicted to be damaging (Polyphen 0.99 Sift 0.00) NM_144769; exon1, c.C517A, p.H173N in the WES data in oncocytoma sample T3. Further screening of the 40 additional oncocytoma samples also identified the same variant in a further three samples, two of Japanese decent and one of German decent, these were confirmed by Sanger sequencing in both directions on the original stock DNA three times. Our collaborator Dr. M. Yao, managed to obtain an additional 4 Japanese oncocytoma samples, which were also screened, one of these samples was also identified to possess the c.C517A, p.H173N variant. Unfortunately, however for the additional samples identified, the matched normal were unavailable; therefore the somatic status is unable to be confirmed (Figure 5-10).

In summary a total of 5 oncocytoma samples were confirmed as possessing SNV c.C517A, totalling 10.2% of oncocytoma samples with a MAF 0.05 within our samples. From the small cohort investigated, 23.5% of Japanese samples (4/17) presented with this variant compare to the 3.1% German (1/32). Fishers exact test identified a significant difference between the frequency of the SNV within Japanese samples compared to samples from a German population (P=0.043), leading us to hypothesise that this variant maybe linked to ethnicity, and could act as a precursor or increase susceptibility for oncocytomas within a Japanese population.

To increase confidence that this mutation is tumour specific and not an uncharacterised SNP, all available matched normals for oncocytomas were screened and identified to be wild type (n=34 (12 Japanese and 22 German).
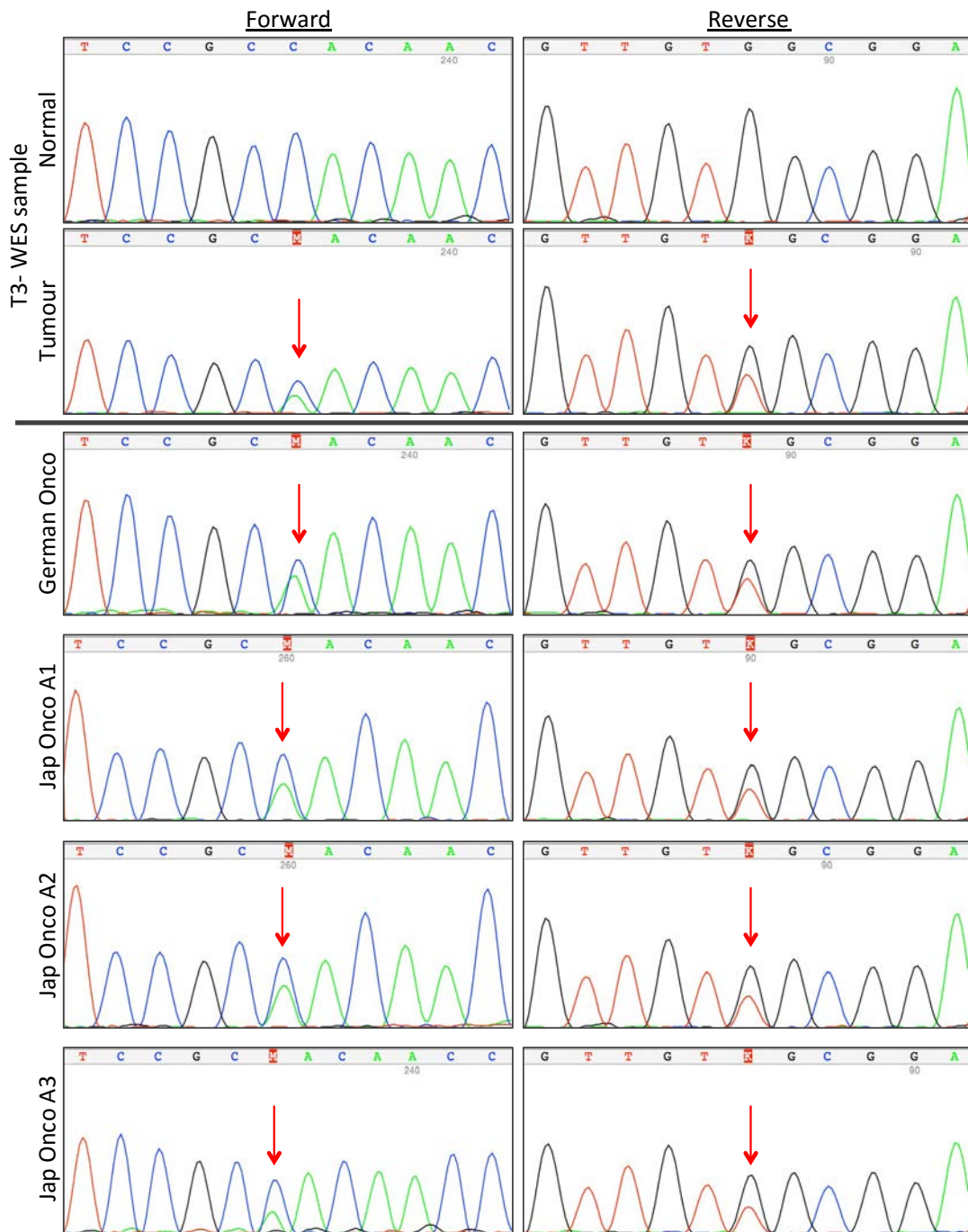
Figure 5-10: Electropherogram of additional *FOXI1* c.C517A, p.H173N SNV
Targeted Sanger sequencing (both directions) of *FOXI1* SNV c.C517A, p.H173N identified in an additional 4 oncocytoma tumour samples (1 of German ancestry and 3 further Japanese ancestry). Repeat of original WES sample T3, further confirms somatic status in this sample, T3 was the only sample with matched normal and thus the only one where somatic status has been confirmed.

A further screen of 101 Japanese N was also undertaken. It should be noted, the normal samples in which the ethnicity was known were from non-tumorous kidney tissue of patients with a renal cancer, the subtype of renal cancers varied with 61 ccRCC, 14 sarcomatoid ccRCC, 14 chRCC, 7 pRCC, 3 Collecting Duct RCC (cdRCC), 1 mutinedus tubular + spindle cell and 1 Angiomyolipoma. These normals shall henceforth be referred to as cancer Normals (cN). All cN were identified to be wt.

To identify if *FOXI1* p.H173N was oncocytoma specific the matched tumours of the aforementioned normals were also sequenced in addition to an extra 14 pRCC, 12 sarcomatoid + ccRCC, 10 chRCC, 5 Angiomyolipoma, 4 cdRCC, 2 Unclassified, 1 Papillary type 2 + Sarcomatoid, 1 AML epithloid type, 1 Juxtaglomerular cell tumours all of Japanese origin (total non-oncocytoma tumours tested n=151).

An ethnicity targeted database search was also undertaken focusing on the Japanese population including a search of the 1000 genome project (105 non-diseased Japanese), the Japanese SNP database (JSNPS; 752 unrelated Japanese samples), and the Integrative Japanese Genome Variation database (1070 Japanese volunteers aged between 20 and 80 years). The SNV c.C517A, p.H173N was not identified in any databases.

When compared to other cancers samples from the TCGA and the published data sets represented on cBioportal, *FOXI1* was most frequently mutated in skin melanoma (6.6% 8/121), while gene application was frequently reported in the TCGA ccRCC 17.3% (72/415) and breast cancer 10.3% (3/29) subsets (Figure 5-11). When analysing the mutation prevalence, was the most frequently mutated although no cancers were reported to contain our SNV of interest, nor any variant mutated to such high a

frequency (Figure 5-11 Figure 5-12a-b). In cosmic, skin cancers also presented with the most frequent mutation rate, of 3.47% (30/864). Once again no other cancer presented with our SNV of interest. Investigation into the location of our SNV, identified it to be located in exon 1 and present in both transcripts of *FOXI1* (Figure 5-12c). Our SNV was located within the highly conserved and functional fork head domain (Figure 5-12b), with our particular amino acid being conserved across multiple species (Figure 5-12d).

Furthermore, expression profiling of *FOXI1* identifies *FOXI1* to be expressed in the kidney ((Figure 5-13a). String db network analysis of human *FOXI1* (Figure 5-13b) predicted two functional partners *PAX8* and *Wnt1* with high confidence of integration (0.924 and 0.715 respectively) that are associated with known KEGG cancer pathways with *Wnt1* also associated with the Hippo pathway (though n=2 and n=1 means that the association of *FOXI1* within the cancer pathways or Hippo pathway was not deemed to be significant). Another 5 other genes (*SLC26A4, TG, DUOXA2, DUOX2* and *PAX8*) within the network were associated with Thyroid hormone synthesis (P = $5.289e^{-8}$) highlighting the complexity of *FOXI1* interactions and network.
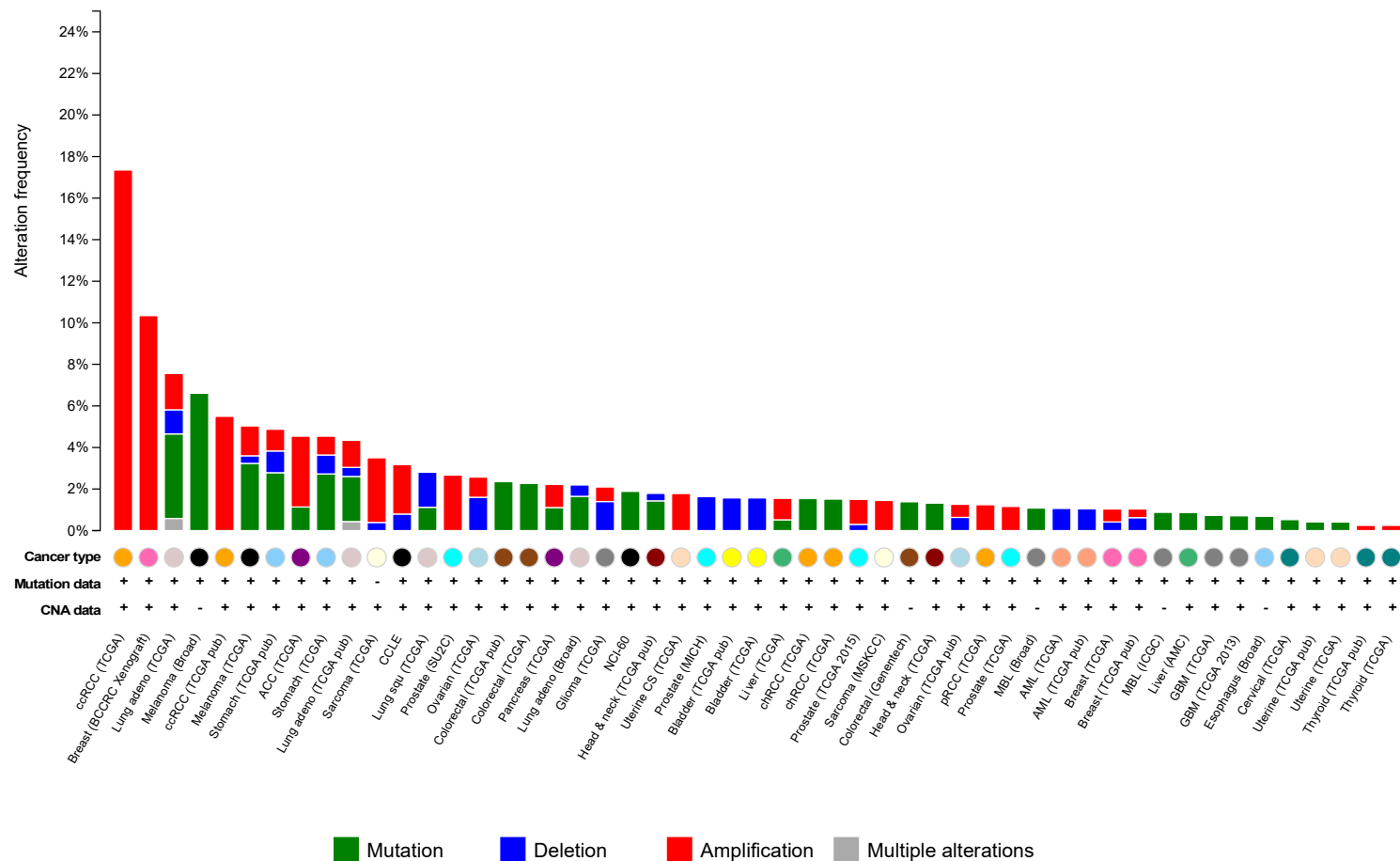
**Figure 5-11: *FOXI1* variations across multiple cancers from Cbioportal**

The mutational and variation landscape of *FOXI1* across different cancers. *FOXI1* in ccRCC and breast cancer was frequently amplified. With gain of copy number in 17.3% (72/415) and 10.3% (3/29) of cases respectively. Skin Melanoma was the most frequently mutated 6.6% (8/121) of cases. Data from TCGA and other publications available cbioportal.org.
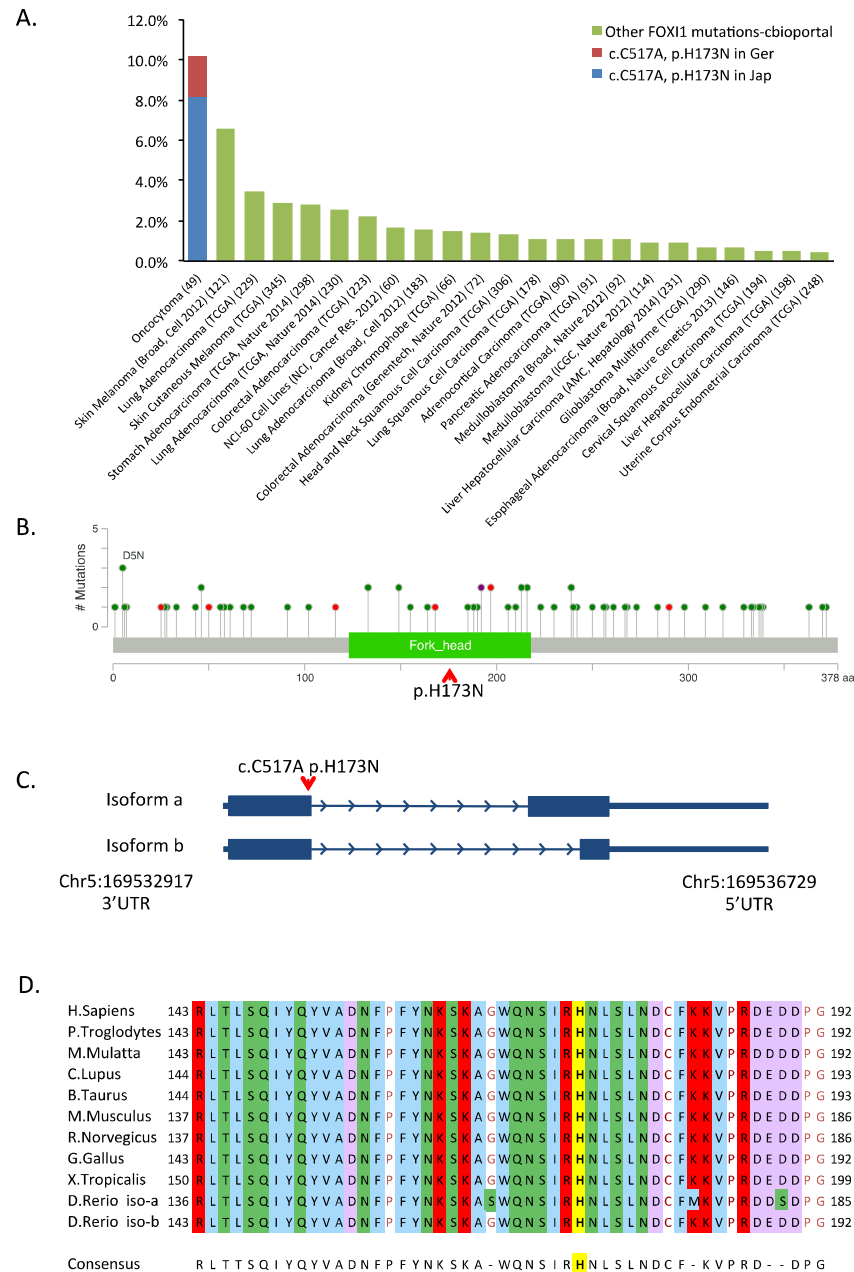
Figure 5-12 *FOXI1* c.C517A, p.H173N characterisation
A) *FOXI1* mutation frequency from in–house renal oncocytoma vs. other cancers from cbioportal.org. The occurrence of SNV c.C517A, p.H173N in oncocytoma (10.2%) was greater than all mutations observed in other cancers. Sample size is show in parenthesis. **B)** Schematic of FOXI1 functional element and the location and frequency of all mutations and cancers from cbioportal.org. Green= missense mutation, red=truncating mutation, purple= multiple mutations. Red arrow, indicates location of the SNV, c.C517A, p.H173N. **C)** Schematic of the genomic structure of FOXI1. Red arrow depicts c.C517A,p.H173N location, present on both transcripts. **D)** Evolutionary conservation of *FOXI1* across multiple species. Amino acid (aa) affected by SNV c.C517A, p.H173N, is highlighted in yellow and bold. aa173 H and the surrounding aa are highly conserved.
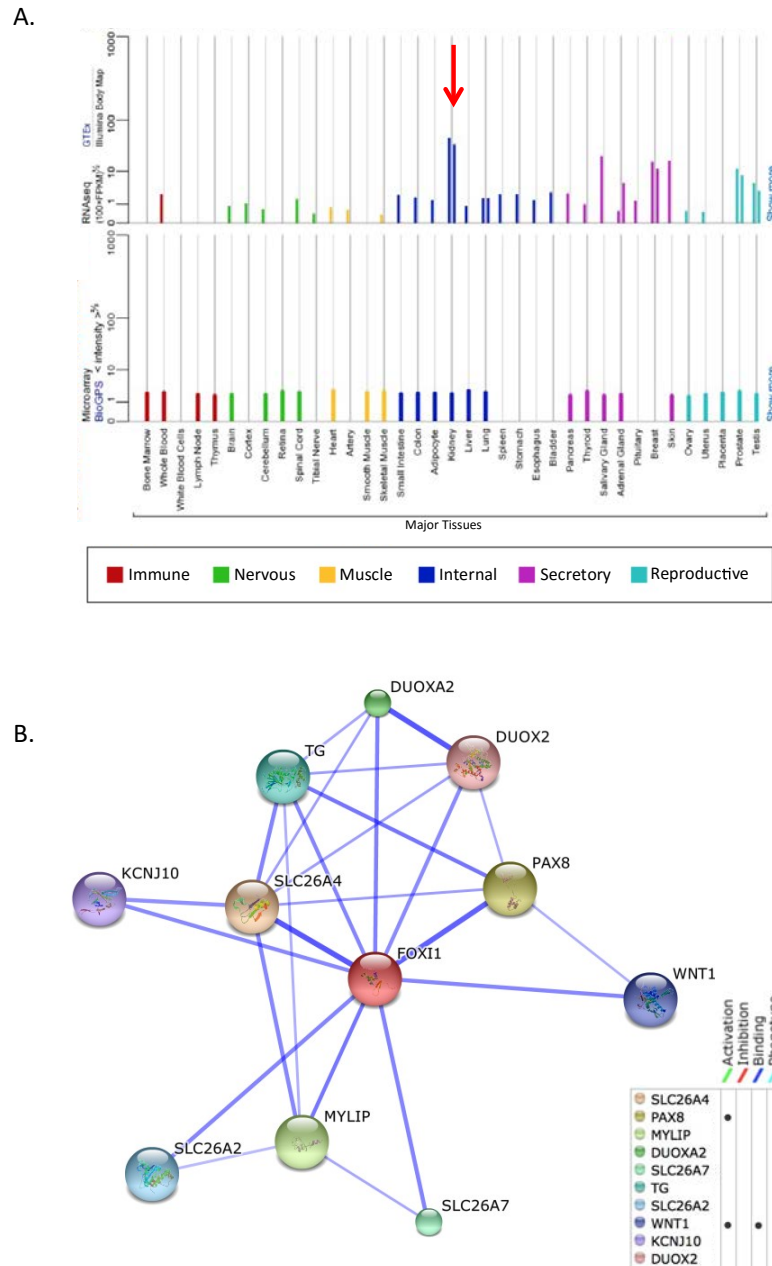
Figure 5-13: Biological characteristics of FOXI1
A) *FOXI1* mRNA expression in human tissue. Characterised by RNAseq and Microarray adapted from Genecards.org. *FOXI1* mRNA expression via RNA-Seq was identified to be particularly increased in kidney (Red arrow). **B)** String-db prediction of human FOXI1 interaction network. Blue lines indicate confidence of interaction, with probability of an interaction documented in 'score' column, high confidence is classified as a score >0.7 Legend also details predicted biological interaction.

5.2.6 TERT promoter

To investigate whether the recently identified TERT promoter hotspots C228T or C250T were associated with oncocytoma, targeted Sanger sequencing of the *TERT* promoter hotspots was conducted in 45 oncocytoma samples. Although neither hotspot variant was detected this result is not wholly unexpected as the hotspot mutations have been associated with more aggressive cancers, rather than a benign tumour such as oncocytoma. However two alternative known SNPs were identified firstly C>T SNV at -348 bp upstream of TSS (chr 5:1295373, rs35226131; global MAF 0.01 (T: 61/C: 4947 alleles)) found in 4.4% oncocytoma samples tested (2/45 with a MAF: 0.02 within our samples). The second SNP identified was A>G -245bp upstream of TERT TSS (chr 5:1295349, rs2853669, MAF 0.3 (G: 1492/ A: 3516)) was found in 44.4% of oncocytoma samples tested (20/45 with a MAF: 0.25 within our samples). Interestingly the presence of rs2853669 is predicted to disrupt a pre-existing Est2 binding site; Hosen et al., 2015 report worse survival in ccRCC patients with the classical hotspot mutations but lacking rs2853996, compared to those possessing the hotspot mutations and the rs2853996 (though the cohort of hotspot mutated samples with and with out rs2853996 was too small for conclusive stats to be preformed). However there is a suggestion that rs2853996 may have a regulatory function in modulating *TERT* activity within cancer, and added another level of complexity into the importance to the *TERT* promoter in malignant disease.

Investigation into structural variation of the *TERT* promoter within oncocytoma was not conducted, due to the exome sequencing not interrogating and identifying the types of variations present (if any) in oncocytoma at the breakpoint region identified with the chRCC study (Davis et al., 2014; Rathmell et al., 2015).

5.3 <u>Discussion</u>

The advances and applications of NGS have provided researchers with a high throughput, in-depth and informative insight into the genetic continuants of many types of cancers and diseases (Rabbani et al., 2014a). WGS can provide additional information into non-coding intronic regions, enhancer regions and identify structural variations and break points; however, as yet WGS remains expensive and requires huge volumes of data processing and bioinformatic analysis with increasingly complex computational pipelines, in order to draw meaningful conclusions. As such it is a technology beyond the capacity both financially and technologically of many laboratories. WES however provides a generally more affordable and accessible alternative. By interrogating the sequence of the exons, at a deeper coverage than WGS, (95% of all exons are covered) (Ku et al., 2012; Rabbani et al., 2014a), WES provides detained information on human exome, corresponding to around 1.5% of the whole genome (Koboldt et al., 2012; Meynert et al., 2014). However, it has been reported that around 85% of disease causing variants are located within the exons (Rabbani et al., 2014a), thus WES is able to identify numerous protein coding variants, and ascertain the biological relevance due to identifying the disruption to the protein sequences. Although WES also generates large data files the release of more and more computational pipelines and bioinformatic techniques to mine the data for somatic variants, LOH and somatic copy number alterations such as Varsan2, MutSigCV (Koboldt et al., 2012; Lawrence et al., 2013) and other bioinformatic techniques; increases the accessibility and reduces the data processing time, promoting a flood of information on somatic mutations in a multitude of cancers and disease such as: small-cell lung cancer (Peifer et al., 2012), ccRCC (Varela et al., 2011), chRCC (Davis et al., 2014), breast cancer

(Fackler et al., 2011), glioblastoma ("TCGA," 2008), and hepatocellular carcinoma (Totoki et al., 2014).

5.3.1 Summary of findings

The focus of this study was to employ WES to identify somatic mutations that may drive the formation of the rare benign renal tumour, oncocytoma. Following a manual bioinformatic data processing pipeline described, a total of 18 variants were identified and confirmed via Sanger sequencing to be somatic, 15 of which were predicted to be damaging in at least one of the *in-silico* prediction software utilised. Eleven of these genes were screened via Sanger sequencing in a cohort of oncocytoma samples totalling 45 samples, with the aim of identifying further recurrent somatic mutations that may attribute to the progression and development of oncocytoma. Of these screened genes *FOXI1,* was identified to possess a recurrent variant, c.C517A p.H173N; first identified within the exome sequencing. This SNV was observed to be more frequent in tumours from a Japanese cohort, 23.5% of samples compared to 3.1% of samples with German ancestry. Further in-house investigation of oncocytoma normals, Japanese cN and other Japanese non-oncocytoma renal tumours did not identify any recurrence, suggesting that the SNV identified is oncocytoma specific. Furthermore studies of publicly accessible databases, with particular focus upon ethnicity (Japanese) revealed no reports of the aforementioned SNV.

The frequency of c.C517A, p.H173N observed, within renal oncocytoma samples suggests this SNV may be an oncocytoma unique hotspot mutation, which could potentially be involved in driving tumourigenesis and inferring a growth advantage (Pon and Marra, 2015; Tamborero et al., 2014; Vogelstein et al., 2013). However, this

variant cannot be confirmed as an oncocytoma specific hotspot mutation until the rest of the *FOXI1* gene is screened in the non-oncocytoma tumours and cN, as alternative pathogenic variants may be identified in the additional samples.

However, the concept of tumour specific hotspot variants is not unique, with Weinreb *et al* and Ueno-Yokohata *et al* also reporting unique tumour specific hotspot variants and alterations. Weinreb *et al* 2014 conducted a study into polymorphous low-grade adenocarcinomas (PLGA) of the salivary glands. They identified a somatic hotspot with two activating missense changes (c.2130A>T or c.2130A>C), that resulted in the same aa change p.Glu710Asp in *PRKD1;* 43 of 59 (72.9%). This variant was associated with increased metastasis free survival and reduced cell migration *in-vitro*. Subsequent database (TCGA and cBioportal) and literature searches led the authors to conclude that p.Glu710Asp in *PRKD1* was a PLGA specific hotspot (Weinreb et al., 2014). Ueno-Yokohata *et al* 2015 investigated paediatric renal tumours, in particular clear cell sarcoma of the kidney (CCSK). They identified that a tandem duplication of between 30-38 aa in exon 16 of *BCOR* in 100% (20/20) of CCSK cases tested, while investigation into a further 193 alternative (non-CCSK) primary paediatric renal tumours identified no duplication or structural abnormality on BCOR exon 16. Therefore proposing tandem repeat in *BCOR* is CCSK unique (Ueno-Yokohata et al., 2015). Characterisation of a tumour specific mutation may therefore provide a useful marker by which the type of tumour can be identified and diagnosed, in addition to providing insights into the biology and pathogenesis of the tumour.

Interestingly, c.C517A, p.H173N was not only identified as an oncocytoma specific SNV, but the variation in frequency of the SNV between the two ethnic populations studied was also interesting, with the SNV being far more frequent within samples

from a Japanese ancestry. It has long been accepted that different ethnicities have different susceptibilities to disease, for example type 2 diabetes is 6 times more likely to occur in someone of south Asian and 3 times more likely in people of African decent (Diabetes.co.uk). According to Cancer Research UK, the incidence of all types of cancers between ethnic groups and gender varies, with the incidence of any type of cancer in white males ranging from 408.2 to 416.8/ 100,000 and similar reports in black males (316.7 to 488.3/100,000), while the incidence of cancer in Asian males was reported to be much lower (168.3 to 258.9/100,000). In females the trend was different with white females having the highest incidence of cancer (from 351.0 to 358.4/ 100,000) compared to and black and Asians women (incidence of 168.4 to 249.8/100,000 and 215.0 to 322.0/100,000 respectively) (Cancer Research UK).

Moreover, studies into the prevalence of mutations within *BRCA1* (associated with increased risk of breast cancer), , have identified a disparity between different ethnic groups (National Cancer Institute). US studies, on an American population showed that Hispanic (3.5% n=393) (significance not declared) (John et al., 2007) and white US population (2.2% (n=508) – 2.9% (n=1145) presented with a higher frequency of the harmful *BRCA1* mutation than black and African-American population (1.3% (n = 341)-1.4% (n=483)) P <0.05 (Malone et al., 2006). Moreover, both studies identified that the prevalence of harmful *BRCA1* mutations was much greater in women of Ashkenazi Jewish descent 10.2% (*n* = 86) P <0.001(Malone et al., 2006), and 8.3% (n=41) (significance not declared) (John et al., 2007) with the majority of mutations were reported to be of one of three key deleterious mutations *185delAG, 5382insC,* and *6174delT.*

Other genome wide association studies and linkage disequilibrium analysis focusing on the association of cancer risk markers between different ethnicities, remarked that characterised risk loci are often poorly linked and not reproducible between ethnic groups. Meaning that where one ethnic population may present with a significant association of a variant with cancer risk, this association may not present as a significant association in another. This suggests a diverse genetic and environmental variation on the association of cancer risk between ethnic groups (Jing et al., 2014).

Large whole exome sequencing investigation by Totoki et al of the mutational landscape of hepatocellular carcinoma (HCC) across different ethnic groups (total of 503 HCC cases +105 TCGA HCC of Japanese, US Asian and European ancestry) identified distinct mutational signatures between, not only ethnic groups but also between genders. HCC is reported to be most prevalent in East Asian and African population, but the incidence is increasing in western populations, with many etiological cofactors being associated, however the finding of Totoki et al also suggest a mutational process that is associated to ancestry, further highlighting the variation and complexity ethnicity can present, even when studying the same disease (Totoki et al., 2014). No association between *FOXI1* c.C517A p.H173N SNV occurrence and gender was noticed.

FOXI1 is a member of the Fork head binding proteins, a family of at least 43 DNA binding proteins, subdivided into smaller groups, although all FOX proteins are characterised by a highly conserved 100aa DNA binding domain that form a winged-helix domain, with many FOX proteins having been associated with a multitude of

cancers (Katoh and Katoh, 2004). The most extensively studied of the FOX family appear to be *FOXM* and *FOXO* subgroups.

FOXM has been summarised to be amplified in 5.6% of breast cancers and 42% of non-Hodgkin's lymphomas, (particularly; diffuse large B-cell lymphoma, follicular lymphoma, and B- cell chronic lymphocytic leukaemia) and unregulated in a plethora of cancers including lung, head and neck squamous cell carcinoma and bladder cancer, among others. *FOXM1,* transcription is induced by hedgehog signalling, and *FOXM1* is self activated by the RAS signalling pathway, transcribing growth associated factors and promoting malignant phenotypes (Katoh et al., 2013). Thus overexpression has been suggested to be a biomarker for poor prognosis in non-small cell lung cancer (J. Zhang et al., 2015)

In the absence of insulin or growth factors, the FOXO family transcribes genes promoting cell cycle arrest, cell differentiation, stress resistance, or apoptosis; determining cell fate under different external conditions and factors and different binding partners. Thus FOXO's have been described to act as tumour suppressors (Greer and Brunet, 2005). For example, upon interacting with SMAD, FOXO transcribes p21, resulting in cell cycle arrest, but when associated with p53, FOXO/P53 reduces p53 inhibition of SIRT, which in turn deacetylases FOXO driving the transcription of stress resistance factors and responses (Calnan and Brunet, 2008; Greer and Brunet, 2005; Milan et al., 2015). Activation of the PI3K-AKT signalling pathway by insulin or growth factors (e.g. TGFB) results in the phosphorylation of FOXO (Fu and Tindall, 2008), while FOXO1 and FOXO3, are also phospholated by MST1/2 core members of the Hippo pathway (Pan, 2010). Several of the *FOXO* family have been associated with chromosomal breakpoints and gene

fusions in several diseases, for example *PAX3:FOXO1*(t(2:13)(q35;q14)) or *PAX7:FOXO1* (t(1:13)(p36;q14)) fusions are associated with alveolar rhabdomyosarcomas, while *FOXO3:MLL* t(6;11)(q21;q23) and *FOXO4:MLL* t(X;11)(q13;q23) has been identified in acute leukaemia (Greer and Brunet, 2005; Katoh et al., 2013)

Additionally FOXO1 is often deleted in prostate cancer, while in Ewing sarcoma, binding of the oncogenic EWS-Fli1 fusion to the FOXO1 promoter, results in down regulation of the mRNA (Katoh et al., 2013; Yang et al., 2010).

Point mutations in *FOXA1,* have been associated with prostate cancer, with FOXA1 binding to the dense chromatin, inducing relaxation of the chromatin and allowing transcription (Katoh et al., 2013). *FOXF* has been characterised as a TSG *in-vitro*, in several different cancer cell lines, with knockdown of *FOXF* resulting in increased cancer cell migration and invasion. Moreover it was shown that *FOXF* was up-regulated by p53 (Tamura et al., 2014). Although not all FOX genes have been classified as possessing tumour suppressive activity, FOXG for example is regarded as an oncogene due to its inhibitor effect preventing FOXO binding to the p21 promoter (Calnan and Brunet, 2008; Fu and Tindall, 2008; Katoh et al., 2013).

Little has been investigated into the role of FOXI1 in regards to cancer, however FOXI1 has been identified to bind to condensed chromatin and remains bound even during mitosis, in a similar manner as TATA and GAGA transcription factors, potentially having a role in modulating chromatin structure (Yan et al., 2006). In addition, *FOXI1* has an important role in the complex signalling networks of embryogenesis and organogenesis particularly in the development of the kidney, jaw and inner ear in zebra-fish (Aghaallaei et al., 2007; Yan et al., 2006) and mice. (El-

Dahr et al., 2008). Within the kidney, *FOXI1* has been identified to mediate the terminal differentiation of epithelial cells within the collecting duct into intercalated cells, with *FOXI1* transcription driven by the p53/p73 transcriptional network (El-Dahr et al., 2008; Guo et al., 2015). Furthermore, *FOXI* initiates the transcription of intercalated cell-specific genes, including vacuolar H+-ATPase Proton Pump subunit transcription, particularly the A4 subunit where expression was additionally localised in *FOXI1* positive epithelial cells in the inner ear, kidney and epididymis (Blomqvist et al., 2004; Vidarsson et al., 2009). While, the loss of *FOXI1* has been identified to result in a lack of intercalated cells in renal collecting duct, and distal renal tubular acidosis (Blomqvist et al., 2004; El-Dahr et al., 2008). Thus *FOXI1/p53* expression promotes intercalated cell identity, the cell type from which oncocytoma are derived (Lopez-Beltran et al., 2006) .

Within the inner ear epithelium development, *FOXI1* has also been reported to up regulate oncogenic associated transcription factor *PAX8* (Aghaallaei et al., 2007)*,* a gene that has since been characterised as a biomarker for non-mucinous surface epithelial ovarian carcinomas (Nonaka et al., 2008). However the role of *FOXI1* on cancer genesis has yet to be investigated.

It was hypothesised, that *FOXI1* c.C517A p.H173N is an oncocytoma specific hotspot variant that is primarily associated with a Japanese ancestry. The SNV resides in the highly conserved fox head DNA binding domain, therefore, c.C517A p.H173 may be involved in the formation or susceptibility for developing renal oncocytoma. It is however vital, that before any definitive conclusions can be drawn that more Japanese oncocytoma samples, preferably with matched normal controls are sourced. Furthermore the, whole *FOXI1* gene should sequenced in both non-oncocytoma

tumour samples and cN, this would identify if alternative somatic variants are present in other tumours or whether the variant *FOXI1* c.C517A p.H173N is an oncocytoma specific hotspot variant. At present we have applied to our collaborator for further samples. Additionally the effect of this mutation on cell growth should be investigated *in-vitro*, to establish if any pathogenic effect is observed. This currently being assessed by the generation and transfection of a plasmid construct containing the mutated *FOXI1* open reading frame, with the specific mutation induced *in-vitro* by site directed mutagenesis into a kidney cell line such as HEK293. The effect on growth will be assessed via anchorage independent colony formation assays (section 2.6.8) .

Until *in-vitro* studies investigating the effect of *FOXI1* SNV have been conducted, it is impossible to hypothesise if *FOXI1* possesses tumour suppressor qualities, like other members of the FOX family, such as *FOXF1* or if it is associated with oncogenic functions that are exacerbated or unregulated by the presence of c.C517A, p.H173N.

With the rise of the TCGA, WES and WGS, many genetic profiles of cancers have been analysed, with more emphasis being placed on the rare cancers. Recently multiplatform analysis of chRCC, identified frequent somatic mutations in TP53 32% of cases (33% 21 of the 66 samples) and PTEN (9% 6 of 66 samples), and multiple genomic rearrangements were identified with in the *TERT* promoter (6/50) (Davis et al., 2014). Another WES study into familial RCC, has recently identified *CDKN2B* germline mutations to predispose patients to the disease in around 5% of cases tested (Jafri et al., 2015).

A study into multiple non-clear cell renal carcinoma subtypes, including pRCC (n=67), chRCC (n=49) and oncocytoma (n=35), identified significantly mutated genes unique

to each subtype. Although chRCC and oncocytoma mutation profile clustered together under unsupervised clustering, subtype specific mutation profile was reported including; for pRCC; *MET, SLC5A3, NF2, PNKD, CPQ, LRP2, CHD3, SLC9A3R1, SETD2* and *CRTC1,* while for chRCC; *TP53, PTEN, FAAH2, PDHB, PDXDC1* and *ZNF765* were hypermutable. For the oncocytoma samples *ERCC2 and C2CD4C* were identified to be significantly mutated although no recurrence was reported (Durinck et al., 2015)*.* Furthermore, we did not identify any recurrent somatic mutations when comparing the genes listed in the supplementary data of Durinck et al, to our own data. Additionally Durinck et al. did not report finding any *FOXI1* variants, however, we propose that this might be due to the different ethnicity investigated, our Japanese samples were extracted in Japan from patients with oncocytoma while Durinck et al. investigated 35 oncocytoma, samples obtained from patients in the USA from a wide variety of patients of multiple ancestry groups that were not specified.

5.3.2 Limitation of this study

A significant limitation of WES is the huge volumes of data produced. Raising issues regarding how best to process and filter the data and draw meaningful and relevant conclusions.

However, several computational pipelines have been released to assist with the processing of this large data pool, and to identify and classify the elusive variants driving cancer. It has been widely accepted that finding recurrent mutations in a known pathway is a good starting point for identification of driver mutations, but as more WES and WGS is under taken more and more focus is being turned to identify driver mutations and de-novo variants and genes previously missed through targeted analysis (Raphael et al., 2014; Tamborero et al., 2014).

For each programme the filtration process differs, and is therefore likely to produce discrepancies within the output list (Kim and Speed, 2013). Although a manual approach to the data processing was ultimately undertaken in this study, Varscan2 analysis was also conducted. A computational pipeline revolving around statistical comparison of read depth and variant call rate. Reassuringly 94% of the variants identified as somatic (by the manual pipeline) were also present in the Varscan2 output. However there were more variants predicted as somatic by Varscan2, as well as analysis of LOH and copy number alterations, something our manual pipeline did not account for. Posing the issue that, perhaps our filtration method was too rigorous, with variants being discredited that may perhaps have been of interest. It should be noted however that due to the algorithms employed via Varscan2, the pipeline is susceptible to array bias (slight variations between different array run on the same machine, resulting in different read depth and coverage of genes) (Koboldt et al., 2012), something that was observed for tumour normal pair T1/N1, (where the normal was repeated at a later date due to low DNA concentration), as a result Varscan2 predicted >200 somatic variants.

Though not remarked upon within this study, it has been previously observed that a high number of what are termed "fishy" genes are recorded via WES to contain cancer specific mutations, following bioinformatic filtering. These are genes that are highly mutated and seem to be associated with cancer but cannot be explained, such as *TTN* encoding the muscle protein titin, genes encoding olfactory receptors, many of the membrane associated mucin family in particular *MUC16* and *MUC4* and other large proteins. The belief is that the prevalence of these genes within WES data may be due to tumour heterogeneity or due to enrichment bias during the exome enrichment stage of WES (Lawrence et al., 2013; Ledford, 2013). To address these

concerns and the risk of tumour/normal contamination MutSigCV uses an algorithm that identifies somatic mutations while also comparing each variant to the expected background frequency of mutations (by comparing the frequency of protein coding variants to synonymous and calculating the probability that a base is mutated by chance) for each loci. Variants exceeding the pre-calculated mutation threshold are deemed to be significant. Thus accommodating the background mutation frequency, into the analysis one would hope to improve upon identification and characterisation of cancer driver genes and pathways (Evans et al., 2013; Lawrence et al., 2013; Watson et al., 2013),

Though many of the usual suspects were incorporated within our raw exome data, our filtration method resulted in none being classified as somatic within our samples, however identification of these type of genes does raise concern that *MYOM1* and *OBSCN* might also belong with this classification of genes due to their size and association with titin*. Although* the MutSigCV pipeline was not used on our data, in hindsight a combination of manual, Varscan2 and MutSigCV may have yielded a refined gene list with potentially novel driver variants, specific for oncocytoma, or pathways affected by an above background rate of mutation.

Another limitation of WES, is that it only encompasses around 1.5% of the genome, although a large amount of data is still produced for this region, it does mean that, unlike in WGS; regulatory, enhancer, promoter regions such as the TERT promoter (Horn et al., 2013; Rathmell et al., 2015) and structural variations including translocations and inversions, are omitted (Davis et al., 2014; Suzuki et al., 2014). Furthermore, the physical method of WES with exome targeted capture and PCR amplification has been reported to result in an ununiformed and variable read depth,

with lower coverage at exon boundaries, shorter fragments, more prone to false positives due to misalignment to the reference genome particularly in repetitive regions and even a bias during exon enrichment for exons that are less GC-rich. This is possibly one of the reasons for the prevalence of "Fishy" genes within WES projects (Ku et al., 2012; Ledford, 2013; Meienberg et al., 2015; Meynert et al., 2014; Sims et al., 2014). Other reports have also identified variation in the exonic regions captured between different commercially available exome capture kits, meaning that variants and exonic regions may be missed or poorly covered depending on the kit used (Lelieveld et al., 2015; Sims et al., 2014).

Tumour heterogeneity is another challenge faced when conducting high throughput sequencing and analysis, and can dilute the sensitivity of WES and WGS to detect heterozygous mutations. Firstly tumours consist of more than one tissue sub clone, and can possess multiple different genetic abnormalities within each tissue subtype, adding complexity to the analysis to identify driver mutations. Additionally, tumour samples are often contaminated with normal noncancerous tissue, as tumours usually reside alongside the normal healthy tissue, some of which may often be included in the tumour extraction as a buffer to ensure the entire tumour is removed from the patient. In order to combat this, it is vital that a sufficient read depth is obtained for both WES and WGS with modern technologies striving to achieve an average read depth of >100 and >30 respectively (Watson et al., 2013). Original methods to asses tumour purity were conducted visually a by skilled pathologist, however there are some computational programmes being released which can calculate tumour purity at the present these tools are optimised for SNP arrays, although they can be adapted for WES and WGS. Programs such as Varscan2 currently acknowledge tumour heterogeneity as an issue in detecting SNVs and

require a user-entered estimation of tumour purity. (Koboldt et al., 2012; Raphael et al., 2014). Although NGS and genomic investigation is both informative and predictive of the effect that a variant, or copy number alteration may have on a gene and transcription, this not necessarily representative of what is happening at protein level. Particularly because many proteins possesses post translational modifications that effect their function. Ideally to obtain a full representation of the biological profile of different cancers, genomic investigation would be coupled with genome wide proteomic analysis (Akbani et al., 2014).

Regarding experimental design of this study, a major limitation and a steep learning curve revolved around the lack of matched normal tissue originally sent for WES. As observed within this study, we could not identify a logical approach to analyse the vast data from the unpaired samples (T4 and T5), as a result the samples ended up being put aside during further filtration, with our study focusing on the matched T/N pairs. Additionally these samples cold not be analysed via computation pipelines, as they require the matched normal in order to identify cancer specific variants. Another limitation was the small cohort size, firstly the small number of samples exome sequenced and secondly the limited number of samples screened via sanger sequencing and thirdly the limited availability of matched normal required to confirm somatic status. The cohort of Japanese samples included in this study was small totalling only 17 oncocytoma samples screened for *FOXI1*, of which only one sample was able to be confirmed somatic for *FOXI1* p.H573N due to the lack of corresponding normals. However oncocytoma is a rare tumour and to acquire sufficient samples from distinct ethnicities has been challenging and dependent on the good will of our collaborators. By expanding the cohort and WES additional T/N pairs, further recurrent variants or susceptibility markers may be identified.

Having discussed the limitations of the WES technology and analysis and our own limitations regarding the oncocytoma cohort, one final remark is regarding the lost data filtered out for various reasons through the *in-silico* analysis. For example a blanket removal of high frequency SNPs was employed, although the vast majority are likely to be of no interest, some SNPs may possess clinical significance. For example the previously mentioned SNP rs2853669 (MAF 0.30) within the *TERT* promoter was reported to increase survival time in patients presenting with the classical TERT hotspot mutations, C228T or C250T (Hosen et al., 2015), while the presence of SNP rs6983267 (MAF 0.39) on chromosome 8q24 has been widely associated with a predisposition to colorectal and prostate cancer (Tuupanen et al., 2009; Ensemble.org). A classic textbook example within pharmarmacogenetics, shows that the presence of many well characterised SNPs, the most notable being *CYP2C19*2* (c.681G>A) Rs4244285 MAF 0.22 results in patients being a poor drug metabolisers reducing the efficacy of several different drugs including some antidepressants, anti-convulsant (mephenytoin) and anti-platelet drug clopidogrel (Scott, 2011).

Moreover, a recent study by Supek et al. 2014, has highlighted a possible oncogenic effect for the previously discarded synonymous mutations. By analysing the prevalence of synonymous mutations in data obtained from TCGA and COSMIC for 11 different cancer tissues (prostate, breast, renal, blood, lung, colorectal, ovarian, melanoma, head-neck, thyroid and brain) with >200 samples per cancer type, Supek at al identified an increase if the prevalence of synonymous mutation in oncogenes compared to TSG in a cancer/tissue specific manner. Although not altering the protein aa sequence, synonymous mutation may alter or generate exon splice sites, resulting in abnormal splicing of the mRNA. A key example was described in TP53 (a

TSG enriched with synonymous mutations that deviated from the normal finding) showing that a synonymous variant at the 3' region of exon 6 created a splice site that resulted in the production of an augmented mRNA (Supek et al., 2014).

5.3.3 Future work

As previously mentioned future work should focus on further identification of the characteristics of the *FOXI1* SNV c.C517A, p. H173N via *in-vitro* investigation, in order to establish the effect of this SNV on growth and even cell migration. At present a construct containing the specific variant is being generated and the effect of the mutation on cell growth is to be analysed by soft agar colony formation (Currently being conducted by Abdullah Alholle (University of Birmingham) and Dr Mark Morris (University of Wolverhampton). The long term, experimental aim would turn towards proteomic investigation via such techniques as mass spectrometry or co-Immunoprecipitation with the aim of identify binding partners and potential pathways in which *FOXI1* may be involved in.

Further screening of *FOXI1* in a larger cohort of sporadic oncocytoma should also be undertaken and ideally included other ethnicities to confirm our hypothesis, and the rest of the *FOXI1* gene should also be screened in the non-oncocytoma tumour and cN samples to identify if other somatic variants are present in different RCC subtypes. In addition, investigation into structural variation or mRNA splicing abnormalities via RNA-Seq and LOH studies should also be undertaken to get a clearer picture as to the role of *FOXI1* in oncocytoma.

5.3.4 Final summary

The reduction in cost, increased accessibility of facilities and services and advancements in bioinformatic pipelines has enabled WES to be widely utilised in the

314

profiling of the somatic exonic landscape of cancers. Although not without its limitations WES has provided a wave of information and insights into many cancers. Though, the variable read depth and enrichment process fundamental to the technology mean that it is inferior to the more uniform WGS. As the cost of WGS becomes more affordable, and bioinformatic pipeline more user friendly, more and more labs will be turning to WGS. Additionally the continued development of publically accessible databases detailing and characterising enhancer and regulatory regions such as ENCODE (Rosenbloom et al., 2013, Encodeproject.org) will boost data processing and interpretation of intronic region variants. Thus it should be predicted that a fresh wave of cancer genomic information will be published, with novel structural variants, and further information on enhancer and other non-coding regulatory variants. This is the aim of the recently launched Genomics England 100,000 genomes project, where the genomes of 75,000 patients with various rare disease and 25,000 tumours will be sequenced and compared to clinical data with the ambition of creating a new genomic medicine service for the NHS, that can enhance the medical research, diagnosis, treatment and overall care for the patient and the disease (Siva, 2015; Genomicsengland.co.uk).

Chapter Six: **Final conclusion**

As previously mentioned cancer is the leading cause of death worldwide. Although kidney cancer is not the most common cancer in the UK, as with most adult cancers, the strong link with incidence and age raises huge concerns as the population ages. Furthermore as traditional chemotherapy and cytotoxic drug treatment is of limited success in the treatment of RCCs (Linehan and Ricketts, 2014). The most frequently used treatment options include surgery (nephrectomy) or radiotherapy, while pharmaceutical intervention predominately involves targeted therapy with treatments such as the anti-VEGF molecules (sunitinib or pazopanib both Tyrosine-kinase inhibitors) or by mTOR inhibitors such as temsirolimus another TKI (Courthod et al., 2015; Linehan and Ricketts, 2014). However, patients response to treatment is dependent on the subtype of RCC (Bylow et al., 2009) it is therefore vital that an accurate diagnosis is obtained quickly in order to ensure the most efficient course of treatment is applied.

Consequently the aim of this study was to investigate and characterise the genetic and epigenetic profiles of rare renal cancers using modern multi-parallel high throughput arrays and sequencing techniques. The genetic investigation into rare diseases and cancers is vital for the identification of candidate genes and deregulated pathways, that contribute to disease progression. Successful identification and understanding of the genes and pathways playing a role in disease may in turn help to identify novel therapeutic targets which can be exploited either for the treatment of these rare diseases for use as biomarkers in the diagnosis of subtypes. Within this study three approaches were undertaken to carry out these investigations:

Within this study three approaches were undertaken to assess the aim:

Firstly the study focused on differentiating between the global DNA methylomes of chRCC and renal oncocytoma using the Infinium HumanMethylation450 BeadChip (chapter three). chRCC and renal oncocytoma are two histologically similar tumours that possess different courses of progression; for example chRCC is a malignant form of RCC while renal oncocytoma is a benign neoplasm, with different courses of treatment. Therefore acquiring an accurate diagnosis is vital to ensure that the optimal course of therapy is undertaken. Through the analysis of the differential methylation profiles of the two histologies, it was hoped that differential methylated genes could be identified that may potentially then be used as biomarkers to distinguish between the two histologies.

This analysis identified distinct methylation patterns, with regard to both chRCC and renal oncocytoma tumours , which were found to present with increased global hypomethylation when compared to the more aggressive ccRCC and pRCC. However due to the well known associations of hypermethylation at CpG island gene promoter regions within cancers, the main focus of the investigation aimed to identify differentially hypermethylated genes between the two RCC histologies. As having the ability to clearly differentiate between the two histologies would then enable more accurate diagnoses and therefore, enable the specific and tailored management of the tumour types.

This analysis identified 30 genes to be differentially hypermethylated (28 in chRCC and 2 in renal oncocytoma). Further comparison of the DNA methylation with ccRCC and pRCC data from the TCGA network identified two candidate genes *SPG20* and *NPHP4* to be significantly hypermethylated in chRCC compared to the other 3 histologies analysed (ccRCC, pRCC and renal oncocytoma). As previously discussed

SPG20 is a gene involved in cytokinesis and cytokinesis arrest. Methylation of *SPG20* has previously been identified as an early biomarker for some colorectal cancer cases (Lind et al., 2011). *NPHP4* is involved in the development of primary cilium, and has also been associated with the Wnt pathway. However, further investigations to confirm that methylation is responsible for suppression of *SPG20* and *NPHP4* gene expression in chRCC is still required. Additionally functional analysis is also required to ascertain the biological impact of loss of *SPG20* and *NPHP4* on cell growth.

Although not the main focus of this study, differential hypomethylation investigations identified 41 genes to be differentially hypomethylated between the two histologies (36 in chRCC and 5 in renal oncocytoma). However, the majority of the hypomethylated probes were mapped to lone (open sea) CpG loci, and thus require further investigation to determine their biological relevance.

The second section of this study (chapter four) focused upon the investigation of methylation of the RASSF family of genes in advanced sporadic sRCC derived from ccRCC, using a sensitive and quantitative methylation assay known as Methylight. In this study *RASSF10,* was identified to be frequently methylated in sRCC and in late stage ccRCC. Analysis of the clinical data identified a positive correlation between the presence of methylation and the occurrence of metastasis, which suggests that loss of *RASSF10* may be associated with more advanced and aggressive forms of cancer, and therefore may potentially be useful as a biomarker to assess disease progression. It was also identified that methylation of *RASSF10* resulted in a decrease of mRNA, and removal of methylation restored the mRNA expression, with further functional analysis confirming that *RASSF10* acts as a TSG where loss results in increased growth and colony formation. A preliminary mass spectrometry investigation was

conducted to characterise RASSF10 and its associated binding partners under different cell growing conditions with the aim of identifying the pathways in which RASSF10 is involved. From this study we identified that RASSF10 binds to TP53B2 (ASPP2) and possibly even P53 itself when the cell cycle was synchronised, which could suggest a potential role in cell cycle checkpoint regulation. However this study was a preliminary investigation, and the results were not validated, therefore no definitive conclusions may be drawn. In order to be more confident of this hypothesis, the mass spectrometry analysis and cell cycle synchronisation would ideally need to be optimised and repeated. In addition , performing co-Immunoprecipitation studies may help to determine protein binding partners, such as TP53 and RASSF10. It is hypothesised that RASSF10 like other RASSF family members, acts as a scaffold protein, thus co-immunoprecipitation should also be conducted to assess this theory. In conclusion, the observations of this study have led the formation of the hypothesis that RASSF10 is associated with more advanced and aggressive forms of RCC, largely due to loss of the tumours suppressor effect. Loss of RASSF10 has been shown to result in increased cell growth, possibly by missing or reducing cell cycle check point arrest. The same hypothesis was recently drawn by Guo et al., when studying RASSF10 in colorectal cancer.(J. Guo et al., 2015)

Finally, as outlined in chapter 5, WES analysis was undertaken on the benign renal oncocytoma samples in the hope of identifying somatic driver mutations that may be responsible for the development and progression of the tumour. Bioinformatic analysis and in-house confirmation managed to identify 18 somatic mutations. Although further targeted screening of most of these genes did not identify and further recurrent mutations, which suggests that the somatic mutation observed were likely to be passenger mutations with little contribution to cancer progression.

However, a recurrent missense mutation (c. C517A p.H173N), in the *FOXI1* gene and predicted to be damaging by Polyphen2 and SIFT, was found to be present in 10.2% of oncocytoma samples. Analysis of the clinical data revealed a difference in frequency of the variant between different ethnic populations; in particular, the variant was found to be enriched in renal oncocytoma tumours of Japanese individuals (23.5%) compared to Caucasians (3.1%). Subsequent screening of tumour DNA from other Japanese RCC subtypes and DNA from their paired normal tissue did not identify any further observation of the SNV. In addition, searches of publically available databases did not identify any reports of this variant with in ethnically-matched healthy control. Little is known as to the exact biological function of *FOXI1* however it has been found to be associated with the initiation of transcription of intercalated cell-specific genes within the epithelium and kidney development, and is predicted to have a role in modulating chromatin structure (Yan et al., 2006). The expression of *FOXI1* has been shown to be regulated by*ADAM10* and the NOTCH signalling pathway (Q. Guo et al., 2015), as well as by p53/p73 (El-Dahr et al., 2008) Furthermore, *FOXI1* has been associated with directing terminal differentiation of renal epithelial cells within the collecting duct into intercalated cells. Initiating the transcription of intercalated cell-specific genes, such as vacuolar H+-ATPase Proton Pump subunits. The intercalated cells are the same cell type form which renal oncocytoma are derived (El-Dahr et al., 2008; Lopez-Beltran et al., 2006; Vidarsson et al., 2009).

Within this study, the analysis of the TCGA mutation data for multiple cancers via cBioportal was also conducted. *FOXI1* was identified to be frequently amplified in ccRCC, suggesting that *FOXI1* may be overexpressed in renal cancers and may contribute or promote renal tumour formation. This observation coupled with the

high frequency c.C517A p.H173N within renal oncocytomas has led us to the hypothesis that *FOXI1* may act as an oncogene, promoting gene transcription and cellular growth within renal oncocytomas. We also propose that c.C517A p.H173N may be an oncocytoma specific hotspot variant associated with Japanese ancestry.. However functional investigation into the effect of this mutation on cell growth will be required before such conclusions can be drawn.

Future work would entail functional analysis, such as colony formation assays in order to assess how the observations made in this study affect cell growth, and if they contribute to tumourigenesis. In particular, further investigation should be conducted into *SPG20, NPHP4* methylation in chRCC and the *FOXI1* mutation in renal oncocytoma. Characterisation of *RASSF10* by mass spectrometry would require optimising and repeating, and with observations validated by further co-immunoprecipitation, before conclusions may be drawn.

This study has identified that rare renal cancers possess diverse epigenetic and genetic profiles that may be of future use as biomarkers for diagnosis, prognosis or as therapeutic targets. The identification of a recurrent missense mutation within a sporadic renal oncocytoma population is of particularly high interest, with further investigation currently being undertaken by Dr Mark Morris (University of Wolverhampton) and Abdullah Alholle (Clinical and Experimental medicine, University of Birmingham) to assess the biological impact of this mutation on cell growth and whether it may act as an oncogenic driver mutation.

Chapter Seven: **Appendix**

Appendix continues on the following page

Gene	Type	Primer	Primer Sequence 5' - 3'	PCR Program	Annelaing Temp (°C)
ALU	Methylight	F	GGT TAG GTA TAG TGG TTT ATA TTT GTA ATT TTA GTA	Methylight	60
		R	ATT AAC TAA ACT AAT CTT AAA CTC CTA ACC TCA		
		P	VIC - CCT ACC TTA ACC TCC C -MGB		
RASSF1A	Methylight	F	ATT GAG TTG CGG GAG TTG GT	Methylight	60
		R	ACA CGC TCC AAC CGA ATA CG		
		P	CCC TTC CCA ACG CGC CCA		
RASSF2	Methylight	F	GTT CGT CAT CGT TTT TTA GGG	Methylight	60
		R	ACC CTA CGC CCC TCT AAA AC		
		P	TAG GTT TTA GTT TTC GGC GCA		
RASSF6	Methylight	F	FAM - TTA GGA TCG TTG ATC GCG TCG GGG GTA TT	Methylight	60
		R	CCC AAAACA TAA CTC AAC TAA AC		
		P	GAA AAG GAG AAA TAA TTA ATA GTT TTT GG		
RASSF10	Methylight	F	AGCGAGGCATCGCTGCCTAAC	Methylight	60
		R	GGGGCGCTCTCGGCTCAGCACTAC		
		P	FAM-CGGCCCCCGCAGCGCCGAGGC-BHQ		
	CoBRA	Outer F	TTG TTT TTG TTG TTT TYG TYG TTT TAG TAG ATT	Touchdown	55
		Outer R	CRA TTA AAC TTA ACC AAT TTA CRA AAA ACC TTA		
		Nested F	GTG TGG ATT TGT TAG GAA GAG AAG T	Standard	56
		Nested R	CTA TTT CTC CTA AAT CAT AAC CAA CAT AA		
	Expression	F	CCATGACCCAGGAGAAACAG	Touchdown 40 cycles	60-56
		R	TGCTGGCGAATTGTGTGGTC		
GAPDH	Expression	F	GACCCCTTCATTGACCTCA	Touchdown 40 cycles	60-56
		R	CTAAGCAGTTGGTGGTGCAGGA		
SPG20	CoBRA	F	5'AATGYGTAGTTTYGTGAYGTAGTATTATT	Touchdown	56
		Nested R	5'AAAAAACTTTATCTAACACCTATATACACCT3'		
		R	5'AAAACTAACACAAATTAATCTAATACCTATTTAAC3'		
NPHP4	CoBRA	F	TTTYGTYGGGGTTTTTTYGGTATTTATT	Touchdown	56
		Nested f	GTTTTTGTAGGTTTYGTTAGTAGTATTTTTATT		
		R	AAACCACAAACCTAAAAACCRAAAACTAA		

Table 7-1: Methylation and Expression primers and conditions

323

Gene	Primer	Primer Sequence 5' - 3'	PCR Program	Annelaing Temp (°C)	Product size
ADAMTS9	L	TCACCTTGGTTCCATCTG	Std 30	58	376
	R	GGAGTGCCCCAGAATACCTT			
APAF1	L	GTAAAGGGGTCAGAACAGCC	Std 30	58	256
	R	CCTTTGCAGTCTTGTCAGCA			
BCL11B	L	TGCGAGTTCTGCGGCAAG	Std 30	58	363
	R	CTCGTTCTCCAGTAGCAGCT			
CAPN1	L	GACAAGGAGCCTGGAGTCTG	Std 30	58	303
	R	ACAGTCCCAGCACTCTTCTC			
LAMB2_NS1	L	TGAGCTGGTGAGGTTGAAGT	Std 30	58	290
	R	TCTGCTGAGGATGCTACCAC			
LAMB2_NS2	L	CCGTGGCTTCTCAGGAATCT	Std 30	58	304
	R	GCCCCATCCCACCATGTATT			
RASGRF2	L	TTTGGCCCAGTCCTTCATCT	Std 30	58	287
	R	TGCCGAGTGTCTCATCAAGA			
SETD2	L	CTCCAAACCCTCCAGCTGTA	Std 30	58	339
	R	CCACTGGCTCTGTTACTGGT			
SMAD4	L	CAGATAGCATCAGGGCCTCA	Std 30	58	349
	R	CAGAAAACAAAGCCCTACCA			
FLNB	L	AGCCCTCCAAGTCATCAACA	Std 30	58	269
	R	TTACCTGTCACCTTGGCCTT			

Table 7-2: Sequencing primers for analysing variants identified by WES in know cancer associated pathway genes and gene on chromosome 3p.

Gene	Exon	Forward Primer	Reverse Primer	PCR Reagent	PCR Program	Annelaing Temp (°C)	Prod Size
NOLC1	6	CATGGTCCTGACTTGCCCTA	TAGGGGTGGTAGCTGCTTTC	Roche	Touch Down 35 cycles	64-60	354
	7	CAGGGTGTGATGTGTGTGTG	TGGATCACTAGTGCTCGGAG	Roche	Touch Down 35 cycles	64-60	361
	8	TCTGGCCCATACTCAAGTGA	TGCTGCTTTCTTTGCTGGAG	Roche	Touch Down 35 cycles	64-60	536
	9+10A	AAGCCTTCCATCCTTCGGTT	TGTCAGCCTTTCTCGTCAGA	Roche	Touch Down 35 cycles	64-60	525
	10B	CTGTCACCACCAAGTCACCT	GCCCTTGAGCTTCTCTTCCT	Roche	Touch Down 35 cycles	64-60	431
	10C	ACAGAAGGTAGCAGGAGGTG	TGGGTGACACAGTGAGACTC	Roche	Touch Down 35 cycles	64-60	442
	11+12	AAGCGATCCTCCTACCTTGG	CCATCCTCCCTCCCAAGAAT	Roche	Touch Down 35 cycles	64-60	500
	13	TGGTTCAGGTTGGTGGGAAT	TCTGTCCTCATCAAGACCCTC	Roche	Touch Down 35 cycles	64-60	389
DCAF4L2	1a	GGACAGAACTTCCCCTCAGG	AAGTGATTCAGTGAGGCCCA	Roche	Touch Down 35 cycles	64-60	472
	1b	GCACAAAACCCTCTACGTCC	TTCCACAGCGCAGATCAATG	Roche	Touch Down 35 cycles	64-60	431
	1c	TGCTGTTTAATGGCTGTCGC	CGAGAAGAGAAGGCCACACT	Roche	Touch Down 35 cycles	64-60	400
	1d	CACAACCATACCCTCCCCAT	AACAGCACCTTACCCCTAGC	Roche	Touch Down 35 cycles	64-60	285
CNBD2 (C20orf152)	1	GGTAGGAGGAGTGGAGTGGA	GCCATGCTGTGACTCCAAAA	Roche	Touch Down 35 cycles	64-60	332
	2	CCTTGTTCAGCACCATCCAG	CCATTGGGTCTCTCATCTGC	Roche	Touch Down 35 cycles	64-60	345
	3	TGCAAAGTTCTAGGGGCAGT	GTTTCTTCCAAGCCAGGGTG	Roche	Touch Down 35 cycles	64-60	180
	4	GGGAACGGGAGAGTGTACAA	CAGGCTCTTTTGCTCAGTCC	Roche	Touch Down 35 cycles	64-60	3224
	5	AAGAGCACAACCAGTCCAGA	GCATTTGACAGGTAAGGGGC	Roche	Touch Down 35 cycles	64-60	387
	6	AAGATGGAGGCACGGAAGAT	CCCAGACAGAGAGAACAGCA	Roche	Touch Down 35 cycles	64-60	354
	7	CAGCGAAATCTTCCACCCAG	GGTTGCAGAAAAGGGCTCAA	Roche	Touch Down 35 cycles	64-60	361
	8	TCTTCAAAGCCTGGCCTACC	AAGGCTCTATGTGCTGGGAA	Roche	Touch Down 35 cycles	64-60	260
	9	TTTCCCCATGTGCTAGACCA	CATCCCTGACATTCTGCCAT	Roche	Touch Down 35 cycles	64-60	379
	10	CATGTGCAGAAGAGTGAGGG	CTTAACATTCCAGCCCATTCATG	Roche	Touch Down 35 cycles	64-60	273
	11	GAATGAAAATCTGCGGGCA	TCATTTCACCCTTGTCAGCA	Roche	Touch Down 35 cycles	64-60	317
	12	AAGGGTGCAGAAGGAGAGTC	CCTTCCTCCAAGTCCCACAG	Roche	Touch Down 35 cycles	64-60	534
NPBWR2	1a	CCCCAACCAGATACCCAACT	GCTCATCACGGCTAGGAAGT	Roche	Touch Down 35 cycles	64-60	490
	1b	GGTGACCAACGTGTTCATCC	TGTAGACACGGCTGGCCTT	Roche	Touch Down 35 cycles	64-60	425
	1c	CCTTCTTCTCTTTCGCTGGC	GGGCATGATGATGATGGGC	Roche	Touch Down 35 cycles	64-60	542
MYOM1	2	CAAGCATGAACCACCGCAC	AGGGTTACTCAGAGCACACG	Biomix Red	Std 30 cycles	60	499
	3	TCCCAAGTTAAAACTTCCTCTATCG	CAAGTGCTCAATAAATGTTAGCTGC	Biomix Red	Std 30 cycles	60	358
	4	CGGATTTTCTTTTGGTGCAGG	ACACACACACACACCCCTTA	Biomix Red	Std 30 cycles	60	528
	5	AGGGGAGAAGGGAGTGCATTT	TGCCTACCCTTCTTCACTCT	Biomix Red	Std 30 cycles	60	**457**
	6	AGAGGAGGGGAAAGCACTTC	CTTTGGGAGGTAGTGTGGGT	Roche	Std 30 cycles	60-56	506
	7+8	AAAGAGCCATGTGTTAGCCA	GCCATATTTTACATAGAGGTGACAC	Roche	Std 30 cycles	60-56	445
	9	TGACCCAGTTAAATCATTTGCTCA	CCACGCCCAGCAGAATATAC	Biomix Red	Std 30 cycles	60	500
	10	ATGTGTCTGCCCCTTATCCC	CGTGAGACATTGCACCCTG	Biomix Red	Std 30 cycles	60	509
	11	GCCAAAGATTGAGCCACTGT	AAGAGGAGGAGAGGAAAGCC	Roche	Std 30 cycles	60	369
	12	AGGACCTGAGGCTTCTTTCC	GGGAGAGAAACAACCTTGCA	Roche	Std 30 cycles	60	502
	13	CCACAGTGCCCTCTACATGA	AATACACACTGCACCCTCCA	Roche	Std 30 cycles	60	305
	14	AGCGTGGGTGATAGAGTGAA	CCTCCTCCCTCACTCCATTG	Roche	Std 30 cycles	60	470
	15	TCCAGCTTGCATTACAGGGA	AGGTTAAGTACAGCCATCGAGT	Roche	Std 30 cycles	60	438
	16	AGCTGTTCTTTTGACATGACGA	CCACCACATCCAGCCAAAA	Roche	Std 30 cycles	60	416
	17	CTGGGAAACACAGCAAGACC	TCCCAGGATGAGCATGAAGT	Roche	Std 30 cycles	60	575
	18	TTTGTTCTGCTTGTGGTCCT	TGTGTACAAGCCAAACCCAT	Roche	Std 30 cycles	60-56	595
	19	AATGCCAAGTTCATCAGCCT	CTTACAAGGTGCTGGGTGTG	Roche	Std 30 cycles	60-56	426
	20	GTCACGGTTTCATCCACACC	GCGATGAGGGCTCTAGGATT	Roche	Std 30 cycles	60	529
	21	GGGACTCAGTGGGCTAATGT	GGAATCTTACCAGGAGCCCA	Roche	Std 30 cycles	60	513
	22	GTATCCCTGGTGCTAGGTCA	TTGGTCCGGGAAGCTTAACT	Roche	Std 30 cycles	60	508
	23	TGTCCACCTGCATGAGTAGT	TTCGATCTCCTGACCTCGTG	Roche	Std 30 cycles	60	554
	24+25	AGCAAGATGAGAGGAGGTGT	AATCCCATCAGAGCTGTCATC	Roche	Std 30 cycles	60	480
	26	CCTCTGGGCTTTGCAAGAAG	AAACTCTCCCACTCCACCAC	Roche	Std 30 cycles	60	393
	27	GGCCTCAAGTAATCCTCCCA	TCGGCCTCTCAAAGTGCT	Roche	Std 30 cycles	60	502
	28	AATGAGCAGGGTTGTCACAA	TCAAAGCTCACACTTAAGTCCT	Roche	Std 30 cycles	58	429
	29	GCGGCAAGAATGGATCAGG	CCCACTTATCATTTTGGCTCAGA	Roche	Std 30 cycles	58	500
	30	TGTCTTCTCACTCCACGTTCA	TCCACACACATCACAAGGTT	Roche	Std 30 cycles	58	403
	31	CAGCAGCAGAAGAAGCAGAA	ATTTCAATCGTCAGCTCGGC	Roche	Std 30 cycles	60	270
	32+33	TGTTTCCATTTCCTTTCAACCAG	CCTCAAATGATCCACCTGCC	Roche	Std 30 cycles	58	466
	34	ATCCCGGTTCTGTGTGTCAT	AGGTGTTGGGATTACAGGCA	Roche	Std 30 cycles	60	446
	35+36	CTGGACCGATCACAGCAATC	ACCACTTCCATTCTCCTCCA	Roche	Std 30 cycles	60	512
	37	TGGGTGACAGAGTGAGACT	TGGAGTACGATGAAGGTGGT	Roche	Std 30 cycles	60	379
	38	CTTATCCTCCCTTTGGTCATGC	AGGGAGGAGAAAGCATGAAGA	Roche	Std 30 cycles	60	491
TERT promoter		CTGGCGTCCCTGCACCCTGG	CTGCCGCTGGCCACGTTCGT	Roche	Std 30 cycles	60	474

Table 7-3: Sequencing primers used for Sanger sequencing screening of candidate genes identified by WES.

Sample	TCGA reference	Sample	TCGA reference
Pap1	TCGA-A4-7286-01A-11D-2137-05	Pap42	TCGA-BQ-7051-01A-12D-1963-05
Pap2	TCGA-A4-7287-01A-11D-2137-05	Pap43	TCGA-BQ-7053-01A-11D-1963-05
Pap3	TCGA-A4-7288-01A-11D-2137-05	Pap44	TCGA-BQ-7055-01A-11D-1963-05
Pap4	TCGA-A4-7583-01A-11D-2137-05	Pap45	TCGA-BQ-7056-01A-11D-1963-05
Pap5	TCGA-A4-7584-01A-11D-2137-05	Pap46	TCGA-BQ-7058-01A-11D-1963-05
Pap6	TCGA-A4-7585-01A-11D-2137-05	Pap47	TCGA-BQ-7059-01A-11D-1963-05
Pap7	TCGA-A4-7732-01A-11D-2137-05	Pap48	TCGA-BQ-7060-01A-11D-1963-05
Pap8	TCGA-A4-7734-01A-11D-2137-05	Pap49	TCGA-BQ-7061-01A-11D-1963-05
Pap9	TCGA-A4-7828-01A-11D-2137-05	Pap50	TCGA-BQ-7062-01A-11D-1963-05
Pap10	TCGA-AL-7173-01A-11D-2137-05	Pap51	TCGA-DW-5560-01A-01D-1590-05
Pap11	TCGA-B1-5398-01A-02D-1590-05	Pap52	TCGA-DW-5561-01A-01D-1590-05
Pap12	TCGA-B1-7332-01A-11D-2137-05	Pap53	TCGA-DW-7834-01A-11D-2137-05
Pap13	TCGA-B9-5155-01A-01D-1590-05	Pap54	TCGA-DW-7836-01A-11D-2137-05
Pap14	TCGA-B9-5156-01A-01D-1590-05	Pap55	TCGA-DW-7837-01A-11D-2137-05
Pap15	TCGA-B9-7268-01A-11D-2137-05	Pap56	TCGA-DW-7838-01A-11D-2137-05
Pap16	TCGA-BQ-5875-01A-11D-1590-05	Pap57	TCGA-DW-7839-01A-11D-2137-05
Pap17	TCGA-BQ-5876-01A-11D-1590-05	Pap58	TCGA-DW-7840-01A-11D-2137-05
Pap18	TCGA-BQ-5877-01A-11D-1590-05	Pap59	TCGA-DW-7841-01A-11D-2137-05
Pap19	TCGA-BQ-5878-01A-11D-1590-05	Pap60	TCGA-DW-7842-01A-11D-2137-05
Pap20	TCGA-BQ-5879-01A-11D-1590-05	Pap61	TCGA-DZ-6131-01A-11D-1963-05
Pap21	TCGA-BQ-5880-01A-11D-1590-05	Pap62	TCGA-DZ-6132-01A-11D-1963-05
Pap22	TCGA-BQ-5881-01A-11D-1590-05	Pap63	TCGA-DZ-6133-01A-11D-1963-05
Pap23	TCGA-BQ-5882-01A-11D-1590-05	Pap64	TCGA-DZ-6134-01A-11D-1963-05
Pap24	TCGA-BQ-5883-01A-11D-1590-05	Pap65	TCGA-DZ-6135-01A-11D-1963-05
Pap25	TCGA-BQ-5884-01A-11D-1590-05	Pap66	TCGA-EV-5901-01A-11D-1590-05
Pap26	TCGA-BQ-5885-01A-11D-1590-05	Pap67	TCGA-EV-5902-01A-11D-1590-05
Pap27	TCGA-BQ-5886-01A-11D-1590-05	Pap68	TCGA-EV-5903-01A-11D-1590-05
Pap28	TCGA-BQ-5887-01A-11D-1963-05	Pap69	TCGA-G7-6789-01A-11D-1963-05
Pap29	TCGA-BQ-5888-01A-11D-1590-05	Pap70	TCGA-G7-6790-01A-11D-1963-05
Pap30	TCGA-BQ-5889-01A-11D-1590-05	Pap71	TCGA-G7-6792-01A-21D-1963-05
Pap31	TCGA-BQ-5890-01A-11D-1590-05	Pap72	TCGA-G7-6793-01A-11D-1963-05
Pap32	TCGA-BQ-5891-01A-11D-1590-05	Pap73	TCGA-G7-6795-01A-11D-1963-05
Pap33	TCGA-BQ-5892-01A-11D-1590-05	Pap74	TCGA-G7-6796-01A-11D-1963-05
Pap34	TCGA-BQ-5893-01A-11D-1590-05	Pap75	TCGA-G7-6797-01A-11D-1963-05
Pap35	TCGA-BQ-5894-01A-11D-1590-05	Pap76	TCGA-GL-6846-01A-11D-1963-05
Pap36	TCGA-BQ-7044-01A-11D-1963-05	Pap77	TCGA-GL-7773-01A-11D-2137-05
Pap37	TCGA-BQ-7045-01A-31D-1963-05	Pap78	TCGA-HE-7128-01A-11D-1963-05
Pap38	TCGA-BQ-7046-01A-11D-1963-05	Pap79	TCGA-HE-7129-01A-11D-1963-05
Pap39	TCGA-BQ-7048-01A-11D-1963-05	Pap80	TCGA-HE-7130-01A-11D-1963-05
Pap40	TCGA-BQ-7049-01A-11D-1963-05	Pap81	TCGA-J7-6720-01A-11D-2137-05
Pap41	TCGA-BQ-7050-01A-11D-1963-05		

Table 7-4: TCGA tumour identification numbers – papillary RCC

Sample	TCGA reference	Sample	TCGA reference
ccRCC1	TCGA-A3-3358-01A-01D-1536-05	ccRCC51	TCGA-BP-4795-01A-02D-1418-05
ccRCC2	TCGA-A3-3367-01A-02D-1418-05	ccRCC52	TCGA-BP-4801-01A-02D-1418-05
ccRCC3	TCGA-A3-3373-01A-02D-1418-05	ccRCC53	TCGA-BP-5168-01A-01D-1418-05
ccRCC4	TCGA-A3-3385-01A-02D-1418-05	ccRCC54	TCGA-BP-5169-01A-01D-1424-05
ccRCC5	TCGA-A3-3387-01A-01D-1536-05	ccRCC55	TCGA-BP-5170-01A-01D-1424-05
ccRCC6	TCGA-AK-3428-01A-02D-1275-05	ccRCC56	TCGA-BP-5182-01A-01D-1424-05
ccRCC7	TCGA-AK-3434-01A-02D-1275-05	ccRCC57	TCGA-BP-5183-01A-01D-1424-05
ccRCC8	TCGA-AK-3445-01A-02D-1275-05	ccRCC58	TCGA-BP-5187-01A-01D-1424-05
ccRCC9	TCGA-B0-4693-01A-01D-1275-05	ccRCC59	TCGA-BP-5190-01A-01D-1424-05
ccRCC10	TCGA-B0-4698-01A-01D-1500-05	ccRCC60	TCGA-BP-5191-01A-01D-1424-05
ccRCC11	TCGA-B0-4701-01A-01D-1275-05	ccRCC61	TCGA-BP-5192-01A-01D-1424-05
ccRCC12	TCGA-B0-4707-01A-01D-1275-05	ccRCC62	TCGA-BP-5194-01A-02D-1424-05
ccRCC13	TCGA-B0-4710-01A-01D-1500-05	ccRCC63	TCGA-BP-5196-01A-01D-1424-05
ccRCC14	TCGA-B0-4821-01A-01D-1500-05	ccRCC64	TCGA-BP-5200-01A-01D-1424-05
ccRCC15	TCGA-B0-4827-01A-02D-1418-05	ccRCC65	TCGA-BP-5202-01A-02D-1424-05
ccRCC16	TCGA-B0-4841-01A-01D-1275-05	ccRCC66	TCGA-CJ-4897-01A-03D-1424-05
ccRCC17	TCGA-B0-4842-01A-02D-1418-05	ccRCC67	TCGA-CJ-5681-01A-11D-1536-05
ccRCC18	TCGA-B0-4845-01A-01D-1275-05	ccRCC68	TCGA-CJ-5682-01A-11D-1536-05
ccRCC19	TCGA-B0-4847-01A-01D-1275-05	ccRCC69	TCGA-CJ-5683-01A-11D-1536-05
ccRCC20	TCGA-B0-4848-01A-01D-1275-05	ccRCC70	TCGA-CJ-5684-01A-11D-1536-05
ccRCC21	TCGA-B0-4849-01A-01D-1275-05	ccRCC71	TCGA-CJ-5686-01A-11D-1670-05
ccRCC22	TCGA-B0-5080-01A-01D-1500-05	ccRCC72	TCGA-CJ-6027-01A-11D-1670-05
ccRCC23	TCGA-B0-5097-01A-01D-1418-05	ccRCC73	TCGA-CJ-6028-01A-11D-1670-05
ccRCC24	TCGA-B0-5099-01A-01D-1418-05	ccRCC74	TCGA-CW-5580-01A-01D-1670-05
ccRCC25	TCGA-B0-5107-01A-01D-1418-05	ccRCC75	TCGA-CW-5588-01A-01D-1536-05
ccRCC26	TCGA-B0-5110-01A-01D-1418-05	ccRCC76	TCGA-CW-5590-01A-01D-1536-05
ccRCC27	TCGA-B0-5113-01A-01D-1418-05	ccRCC77	TCGA-CW-6090-01A-11D-1670-05
ccRCC28	TCGA-B0-5116-01A-02D-1418-05	ccRCC78	TCGA-CW-6097-01A-11D-1670-05
ccRCC29	TCGA-B0-5119-01A-02D-1418-05	ccRCC79	TCGA-CZ-4853-01A-01D-1424-05
ccRCC30	TCGA-B0-5120-01A-01D-1418-05	ccRCC80	TCGA-CZ-4856-01A-02D-1424-05
ccRCC31	TCGA-B0-5121-01A-02D-1418-05	ccRCC81	TCGA-CZ-4859-01A-02D-1424-05
ccRCC32	TCGA-B0-5399-01A-01D-1500-05	ccRCC82	TCGA-CZ-4863-01A-01D-1500-05
ccRCC33	TCGA-B0-5690-01A-11D-1536-05	ccRCC83	TCGA-CZ-4865-01A-02D-1500-05
ccRCC34	TCGA-B0-5692-01A-11D-1536-05	ccRCC84	TCGA-CZ-4866-01A-01D-1500-05
ccRCC35	TCGA-B0-5694-01A-11D-1536-05	ccRCC85	TCGA-CZ-5451-01A-01D-1500-05
ccRCC36	TCGA-B0-5697-01A-11D-1536-05	ccRCC86	TCGA-CZ-5454-01A-01D-1500-05
ccRCC37	TCGA-B0-5698-01A-11D-1670-05	ccRCC87	TCGA-CZ-5455-01A-01D-1500-05
ccRCC38	TCGA-B0-5706-01A-11D-1536-05	ccRCC88	TCGA-CZ-5457-01A-01D-1500-05
ccRCC39	TCGA-B0-5712-01A-11D-1670-05	ccRCC89	TCGA-CZ-5458-01A-01D-1500-05
ccRCC40	TCGA-B0-5713-01A-11D-1670-05	ccRCC90	TCGA-CZ-5462-01A-01D-1500-05
ccRCC41	TCGA-B2-5633-01A-01D-1536-05	ccRCC91	TCGA-CZ-5465-01A-01D-1500-05
ccRCC42	TCGA-B2-5639-01A-01D-1536-05	ccRCC92	TCGA-CZ-5466-01A-01D-1500-05
ccRCC43	TCGA-B4-5378-01A-01D-1500-05	ccRCC93	TCGA-CZ-5467-01A-01D-1500-05
ccRCC44	TCGA-B4-5834-01A-11D-1670-05	ccRCC94	TCGA-CZ-5469-01A-01D-1500-05
ccRCC45	TCGA-B4-5836-01A-11D-1670-05	ccRCC95	TCGA-DV-5569-01A-01D-1536-05
ccRCC46	TCGA-B4-5843-01A-11D-1670-05	ccRCC96	TCGA-DV-5574-01A-01D-1536-05
ccRCC47	TCGA-B8-4621-01A-01D-1500-05	ccRCC97	TCGA-DV-5576-01A-01D-1536-05
ccRCC48	TCGA-B8-4622-01A-02D-1275-05	ccRCC98	TCGA-EU-5905-01A-11D-1670-05
ccRCC49	TCGA-B8-5162-01A-01D-1418-05	ccRCC99	TCGA-EU-5906-01A-11D-1670-05
ccRCC50	TCGA-B8-5552-01B-11D-1670-05	ccRCC100	TCGA-EU-5907-01A-11D-1670-05

Table 7-5: TCGA tumour identification numbers – clear cell RCC included in chapter 3

Index	TCGA reference	Index	TCGA reference
1	TCGA-A3-3357-01A-02D-1418-05	43	TCGA-B0-4846-01A-01D-1275-05
2	TCGA-A3-3358-01A-01D-1536-05	44	TCGA-B0-4847-01A-01D-1275-05
3	TCGA-A3-3367-01A-02D-1418-05	45	TCGA-B0-4848-01A-01D-1275-05
4	TCGA-A3-3370-01A-02D-1418-05	46	TCGA-B0-4849-01A-01D-1275-05
5	TCGA-A3-3373-01A-02D-1418-05	47	TCGA-B0-4852-01A-01D-1500-05
6	TCGA-A3-3376-01A-02D-1418-05	48	TCGA-B0-4945-01A-01D-1418-05
7	TCGA-A3-3385-01A-02D-1418-05	49	TCGA-B0-5080-01A-01D-1500-05
8	TCGA-B0-4688-01A-01D-1275-05	50	TCGA-B0-5083-01A-02D-1418-05
9	TCGA-B0-4690-01A-01D-1275-05	51	TCGA-B0-5092-01A-01D-1418-05
10	TCGA-B0-4691-01A-01D-1275-05	52	TCGA-B0-5094-01A-01D-1418-05
11	TCGA-B0-4693-01A-01D-1275-05	53	TCGA-B0-5095-01A-01D-1418-05
12	TCGA-B0-4694-01A-01D-1275-05	54	TCGA-B0-5096-01A-01D-1418-05
13	TCGA-B0-4696-01A-01D-1275-05	55	TCGA-B0-5097-01A-01D-1418-05
14	TCGA-B0-4697-01A-01D-1275-05	56	TCGA-B0-5098-01A-01D-1418-05
15	TCGA-B0-4698-01A-01D-1500-05	57	TCGA-B0-5099-01A-01D-1418-05
16	TCGA-B0-4699-01A-01D-1275-05	58	TCGA-B0-5100-01A-01D-1418-05
17	TCGA-B0-4700-01A-02D-1536-05	59	TCGA-B0-5102-01A-01D-1418-05
18	TCGA-B0-4701-01A-01D-1275-05	60	TCGA-B0-5104-01A-01D-1418-05
19	TCGA-B0-4703-01A-01D-1275-05	61	TCGA-B0-5106-01A-01D-1418-05
20	TCGA-B0-4706-01A-01D-1500-05	62	TCGA-B0-5107-01A-01D-1418-05
21	TCGA-B0-4707-01A-01D-1275-05	63	TCGA-B0-5108-01A-01D-1418-05
22	TCGA-B0-4710-01A-01D-1500-05	64	TCGA-B0-5109-01A-02D-1418-05
23	TCGA-B0-4712-01A-01D-1500-05	65	TCGA-B0-5110-01A-01D-1418-05
24	TCGA-B0-4713-01A-01D-1275-05	66	TCGA-B0-5113-01A-01D-1418-05
25	TCGA-B0-4714-01A-01D-1275-05	67	TCGA-B0-5115-01A-01D-1418-05
26	TCGA-B0-4718-01A-01D-1275-05	68	TCGA-B0-5116-01A-02D-1418-05
27	TCGA-B0-4810-01A-01D-1500-05	69	TCGA-B0-5117-01A-01D-1418-05
28	TCGA-B0-4811-01A-01D-1500-05	70	TCGA-B0-5119-01A-02D-1418-05
29	TCGA-B0-4813-01A-01D-1275-05	71	TCGA-B0-5120-01A-01D-1418-05
30	TCGA-B0-4814-01A-01D-1275-05	72	TCGA-B0-5121-01A-02D-1418-05
31	TCGA-B0-4815-01A-01D-1500-05	73	TCGA-B0-5399-01A-01D-1500-05
32	TCGA-B0-4816-01A-01D-1500-05	74	TCGA-B0-5400-01A-01D-1500-05
33	TCGA-B0-4817-01A-01D-1275-05	75	TCGA-B0-5402-01A-01D-1500-05
34	TCGA-B0-4818-01A-01D-1500-05	76	TCGA-B0-5710-01A-11D-1670-05
35	TCGA-B0-4819-01A-01D-1275-05	77	TCGA-B0-5711-01A-11D-1670-05
36	TCGA-B0-4821-01A-01D-1500-05	78	TCGA-B0-5712-01A-11D-1670-05
37	TCGA-B0-4822-01A-01D-1275-05	79	TCGA-B0-5713-01A-11D-1670-05
38	TCGA-B0-4823-01A-02D-1418-05	80	TCGA-BP-4177-01A-02D-1418-05
39	TCGA-B0-4824-01A-01D-1275-05	81	TCGA-BP-4760-01A-02D-1418-05
40	TCGA-B0-4827-01A-02D-1418-05	82	TCGA-BP-4770-01A-01D-1500-05
41	TCGA-B0-4828-01A-01D-1275-05	83	TCGA-BP-4782-01A-02D-1418-05
42	TCGA-B0-4841-01A-01D-1275-05	84	TCGA-BP-4795-01A-02D-1418-05

Index	TCGA reference	Index	TCGA reference
85	TCGA-BP-5168-01A-01D-1418-05	128	TCGA-CZ-4856-01A-02D-1424-05
86	TCGA-BP-5169-01A-01D-1424-05	129	TCGA-CZ-4859-01A-02D-1424-05
87	TCGA-BP-5170-01A-01D-1424-05	130	TCGA-CZ-4863-01A-01D-1500-05
88	TCGA-BP-5173-01A-01D-1424-05	131	TCGA-CZ-4864-01A-01D-1500-05
89	TCGA-BP-5174-01A-01D-1424-05	132	TCGA-CZ-4865-01A-02D-1500-05
90	TCGA-BP-5175-01A-01D-1424-05	133	TCGA-CZ-4866-01A-01D-1500-05
91	TCGA-BP-5176-01A-01D-1424-05	134	TCGA-CZ-5451-01A-01D-1500-05
92	TCGA-BP-5177-01A-01D-1424-05	135	TCGA-CZ-5452-01A-01D-1500-05
93	TCGA-BP-5178-01A-01D-1424-05	136	TCGA-CZ-5453-01A-01D-1500-05
94	TCGA-BP-5180-01A-01D-1424-05	137	TCGA-CZ-5454-01A-01D-1500-05
95	TCGA-BP-5181-01A-01D-1424-05	138	TCGA-CZ-5455-01A-01D-1500-05
96	TCGA-BP-5182-01A-01D-1424-05	139	TCGA-CZ-5456-01A-01D-1500-05
97	TCGA-BP-5183-01A-01D-1424-05	140	TCGA-CZ-5457-01A-01D-1500-05
98	TCGA-BP-5184-01A-01D-1424-05	141	TCGA-CZ-5458-01A-01D-1500-05
99	TCGA-BP-5185-01A-01D-1424-05	142	TCGA-CZ-5459-01A-01D-1500-05
100	TCGA-BP-5186-01A-01D-1424-05	143	TCGA-CZ-5460-01A-01D-1500-05
101	TCGA-BP-5187-01A-01D-1424-05	144	TCGA-CZ-5461-01A-01D-1500-05
102	TCGA-BP-5189-01A-02D-1424-05	145	TCGA-CZ-5462-01A-01D-1500-05
103	TCGA-BP-5190-01A-01D-1424-05	146	TCGA-CZ-5463-01A-01D-1500-05
104	TCGA-BP-5191-01A-01D-1424-05	147	TCGA-CZ-5464-01A-01D-1500-05
105	TCGA-BP-5192-01A-01D-1424-05	148	TCGA-CZ-5465-01A-01D-1500-05
106	TCGA-BP-5194-01A-02D-1424-05	149	TCGA-CZ-5466-01A-01D-1500-05
107	TCGA-BP-5195-01A-02D-1424-05	150	TCGA-CZ-5467-01A-01D-1500-05
108	TCGA-BP-5196-01A-01D-1424-05	151	TCGA-CZ-5468-01A-01D-1500-05
109	TCGA-BP-5198-01A-01D-1424-05	152	TCGA-CZ-5469-01A-01D-1500-05
110	TCGA-BP-5199-01A-01D-1424-05	153	TCGA-CZ-5470-01A-01D-1500-05
111	TCGA-BP-5200-01A-01D-1424-05	154	TCGA-B0-4842-01A-02D-1418-05
112	TCGA-BP-5201-01A-01D-1424-05	155	TCGA-B0-4843-01A-01D-1275-05
113	TCGA-BP-5202-01A-02D-1424-05	156	TCGA-B0-4844-01A-01D-1275-05
114	TCGA-CJ-4869-01A-02D-1424-05	157	TCGA-BP-4801-01A-02D-1418-05
115	TCGA-CJ-4882-01A-02D-1424-05	158	TCGA-BP-4993-01A-02D-1418-05
116	TCGA-CJ-4897-01A-03D-1424-05	159	TCGA-BP-5010-01A-02D-1418-05
117	TCGA-CJ-4901-01A-01D-1424-05	160	TCGA-CJ-4920-01A-01D-1424-05
118	TCGA-CJ-4902-01A-01D-1424-05		
119	TCGA-CJ-4903-01A-01D-1424-05		
120	TCGA-CJ-4904-01A-02D-1424-05		
121	TCGA-CJ-4905-01A-02D-1424-05		
122	TCGA-CJ-4907-01A-01D-1424-05		
123	TCGA-CJ-4908-01A-01D-1424-05		
124	TCGA-CJ-4912-01A-01D-1424-05		
125	TCGA-CJ-4913-01A-01D-1424-05		
126	TCGA-CJ-4916-01A-01D-1424-05		
127	TCGA-CJ-4918-01A-01D-1424-05		

Table 7-6: TCGA identification numbers for ccRCC samples used in chapter 4.
Tumour and matched normal for each sample was downloaded

Hypermethylated probes associated with genes in >30% chromophobe

Gene	Cancer	TargetID	CHR	Prevelance		Frequently Methylated in other Kidney Cancers
ABI3BP	Ch	cg15028548	3	6/20	30.00%	
ACN9	Ch	cg10548492	7	7/20	35.00%	
ADAM5P	Ch	cg14742937	8	7/20	35.00%	
ADAMTS14	Ch	cg04838832	10	9/20	45.00%	
ADAR	Ch	cg13392022	1	6/20	30.00%	
AFF1	Ch	cg17455261	4	8/20	40.00%	
AGAP1	Ch	cg17945789	2	11/20	55.00%	
ALX1	Ch	cg13916459	12	6/20	30.00%	
AMH	Ch	cg05345154	19	6/20	30.00%	
ARHGAP10	Ch	cg11713658	4	7/20	35.00%	
ARHGAP25	Ch	cg12093662	2	6/20	30.00%	
ASCL2	Ch	cg10290276	11	6/20	30.00%	Arai E et al 2012 PMID: 22610075
ATP6V1B1	Ch	cg04751811	2	11/20	55.00%	
ATP6V1B1	Ch	cg06288697	2	15/20	75.00%	
BAG2	Ch	cg10230427	6	7/20	35.00%	
BAG2	Ch	cg27164797	6	6/20	30.00%	
BTNL9	Ch	cg16049690	5	6/20	30.00%	
C16orf68	Ch	cg04616793	16	7/20	35.00%	
C17orf101	Ch	cg21525032	17	7/20	35.00%	
C17orf95	Ch	cg27000944	17	7/20	35.00%	
C1orf14	Ch	cg08146495	1	6/20	30.00%	
C20orf54	Ch	cg11959399	20	6/20	30.00%	
C22orf27	Ch	cg08837037	22	6/20	30.00%	
C2CD4D	Ch	cg04296699	1	6/20	30.00%	
C2CD4D	Ch	cg15015892	1	7/20	35.00%	
C2orf43	Ch	cg14757738	2	10/20	50.00%	
C6orf27	Ch	cg13187827	6	7/20	35.00%	
C7orf58	Ch	cg01597480	7	8/20	40.00%	
CACNA1A	Ch	cg11660879	19	6/20	30.00%	
CACNG7	Ch	cg21477176	19	6/20	30.00%	
CAPN2	Ch	cg06756211	1	13/20	65.00%	
CCDC140	Ch	cg09486778	2	7/20	35.00%	
CCDC64B	Ch	cg14134732	16	6/20	30.00%	
CCDC69	Ch	cg08317263	5	6/20	30.00%	
CCK	Ch	cg16864658	3	9/20	45.00%	
CDX2	Ch	cg00806704	13	7/20	35.00%	
CHD9	Ch	cg09608652	16	6/20	30.00%	
CHRM2	Ch	cg13102079	7	6/20	30.00%	
CHST8	Ch	cg22335692	19	6/20	30.00%	
CLEC11A	Ch	cg13152535	19	7/20	35.00%	
CMTM3	Ch	cg26560414	16	6/20	30.00%	

Gene	Cancer	TargetID	CHR	Prevelance		Frequently Methylated in other Kidney Cancers
COL4A1	Ch	cg23008352	13	7/20	35.00%	
CPT1A	Ch	cg13786863	11	7/20	35.00%	
CRY2	Ch	cg16006220	11	6/20	30.00%	
CSDAP1	Ch	cg06301178	16	6/20	30.00%	
CSRP2	Ch	cg21484956	12	6/20	30.00%	
CUX1	Ch	cg02856420	7	10/20	50.00%	
CXCL14	Ch	cg04002608	5	9/20	45.00%	
DBC1	Ch	cg03625109	9	7/20	35.00%	
DCHS2	Ch	cg11934170	4	6/20	30.00%	
DENND3	Ch	cg24476153	8	7/20	35.00%	
DLGAP3	Ch	cg00114029	1	6/20	30.00%	
DLGAP3	Ch	cg18852567	1	8/20	40.00%	
DLGAP3	Ch	cg26228266	1	7/20	35.00%	
DLX1	Ch	cg17737681	2	7/20	35.00%	
DLX4	Ch	cg11472521	17	6/20	30.00%	
DPP4	Ch	cg19350270	2	7/20	35.00%	
DQX1	Ch	cg05652569	2	6/20	30.00%	
DRD4	Ch	cg06299284	11	6/20	30.00%	
ELTD1	Ch	cg15084543	1	12/20	60.00%	
EN2	Ch	cg12034383	7	6/20	30.00%	
EPS8L1	Ch	cg08166750	19	8/20	40.00%	
EPS8L1	Ch	cg12020396	19	11/20	55.00%	
EPS8L1	Ch	cg18320766	19	12/20	60.00%	
EPS8L2	Ch	cg08327690	11	9/20	45.00%	
EVX2	Ch	cg07536910	2	7/20	35.00%	
FAM123A	Ch	cg18815647	13	6/20	30.00%	
FAM125A	Ch	cg02217814	19	6/20	30.00%	
FBLL1	Ch	cg13954457	5	8/20	40.00%	
FHL3	Ch	cg24519393	1	7/20	35.00%	
FLJ22536	Ch	cg08233811	6	6/20	30.00%	
FLJ22536	Ch	cg26158950	6	6/20	30.00%	
FLJ37453	Ch	cg25980637	1	6/20	30.00%	
FOXL2	Ch	cg07135614	3	6/20	30.00%	
GCM2	Ch	cg10074727	6	6/20	30.00%	
GDF6	Ch	cg02246426	8	11/20	55.00%	
GIPC2	Ch	cg01074657	1	10/20	50.00%	
GJA3	Ch	cg22900415	13	6/20	30.00%	
GJB2	Ch	cg00347757	13	6/20	30.00%	
GPRIN1	Ch	cg11214757	5	6/20	30.00%	
GRIK2	Ch	cg21635870	6	8/20	40.00%	
GUCY1B3	Ch	cg04040095	4	7/20	35.00%	
GUCY1B3	Ch	cg18129755	4	7/20	35.00%	

Gene	Cancer	TargetID	CHR	Prevelance		Frequently Methylated in other Kidney Cancers
GYPC	Ch	cg04453971	2	7/20	35.00%	
HAPLN1	Ch	cg12199221	5	6/20	30.00%	
HHEX	Ch	cg02185052	10	9/20	45.00%	
HHEX	Ch	cg09721427	10	6/20	30.00%	
HOXA2	Ch	cg02225599	7	8/20	40.00%	
HOXA3	Ch	cg21134232	7	6/20	30.00%	
HOXA9	Ch	cg03217995	7	8/20	40.00%	McRonald.F et al 2009
HOXC10	Ch	cg20402783	12	9/20	45.00%	
HOXC4	Ch	cg19164987	12	7/20	35.00%	
HOXC5	Ch	cg07080050	12	6/20	30.00%	
HSPB9	Ch	cg08880849	17	8/20	40.00%	
HTR2A	Ch	cg15894389	13	6/20	30.00%	
IGF2BP1	Ch	cg02745847	17	7/20	35.00%	
IRX6	Ch	cg01064265	16	7/20	35.00%	
JSRP1	Ch	cg04887494	19	6/20	30.00%	
KCNH3	Ch	cg27382405	12	8/20	40.00%	
KCNJ8	Ch	cg03352106	12	16/20	80.00%	
KCNJ8	Ch	cg16739441	12	9/20	45.00%	
KCNK2	Ch	cg05129325	1	6/20	30.00%	
KCNK2	Ch	cg11200222	1	9/20	45.00%	
KCNQ4	Ch	cg19689322	1	6/20	30.00%	
KRT27	Ch	cg02399249	17	6/20	30.00%	
LASP1	Ch	cg04569429	17	6/20	30.00%	
LBX1	Ch	cg03053579	10	6/20	30.00%	
LHX2	Ch	cg14425564	9	8/20	40.00%	
LHX8	Ch	cg11842415	1	6/20	30.00%	
LMO2	Ch	cg06148118	11	7/20	35.00%	
LOC100130522	Ch	cg12061113	18	7/20	35.00%	
LOC100132111	Ch	cg24051554	1	8/20	40.00%	
LOC645323	Ch	cg03632704	5	7/20	35.00%	
LOC645323	Ch	cg12991050	5	6/20	30.00%	
LOC645323	Ch	cg13982098	5	7/20	35.00%	
LRRC32	Ch	cg01439670	11	6/20	30.00%	
LRRC32	Ch	cg13633560	11	6/20	30.00%	
LYPD5	Ch	cg11898486	19	7/20	35.00%	
MAML2	Ch	cg24088496	11	8/20	40.00%	
MANEAL	Ch	cg00664416	1	6/20	30.00%	
MEOX2	Ch	cg07395354	7	7/20	35.00%	
MKX	Ch	cg08423533	10	7/20	35.00%	
MKX	Ch	cg26298409	10	6/20	30.00%	
MRC2	Ch	cg24368031	17	9/20	45.00%	
MYO3A	Ch	cg02232704	10	6/20	30.00%	

Gene	Cancer	TargetID	CHR	Prevelance		Frequently Methylated in other Kidney Cancers
NCKAP5	Ch	cg02715592	2	7/20	35.00%	
NEUROD1	Ch	cg01897496	2	7/20	35.00%	
NFE2L3	Ch	cg10536999	7	8/20	40.00%	
NKAPL	Ch	cg17384889	6	6/20	30.00%	
NKX1-2	Ch	cg00486352	10	6/20	30.00%	
NKX6-2	Ch	cg11174855	10	7/20	35.00%	Arai E et al 2012 PMID: 22610075
NOS1	Ch	cg17119907	12	7/20	35.00%	
NPHP4	Ch	cg20383686	1	6/20	30.00%	
NR2E1	Ch	cg18569734	6	6/20	30.00%	
OTP	Ch	cg16703762	5	6/20	30.00%	
OTX2	Ch	cg13023623	14	8/20	40.00%	
OTX2	Ch	cg15607672	14	7/20	35.00%	
PACS2	Ch	cg22976533	14	8/20	40.00%	
PARD3	Ch	cg15292765	10	17/20	85.00%	
PAX3	Ch	cg13767755	2	6/20	30.00%	
PAX3	Ch	cg14093610	2	6/20	30.00%	
PAX5	Ch	cg03261800	9	8/20	40.00%	
PCDH17	Ch	cg12432236	13	6/20	30.00%	Costa VL et al 2011 PMID: 21847011
PCDHA12	Ch	cg09852127	5	7/20	35.00%	
PCDHA7	Ch	cg02051771	5	6/20	30.00%	
PCDHB15	Ch	cg10757144	5	7/20	35.00%	
PCDHB17	Ch	cg21777188	5	6/20	30.00%	
PCDHB3	Ch	cg24586758	5	10/20	50.00%	
PCDHGA2	Ch	cg18781988	5	8/20	40.00%	
PCDHGA4	Ch	cg06757585	5	7/20	35.00%	
PCSK6	Ch	cg07324702	15	14/20	70.00%	
PDE4D	Ch	cg03323696	5	6/20	30.00%	
PHACTR1	Ch	cg13246235	6	9/20	45.00%	
PHOX2B	Ch	cg01416712	4	8/20	40.00%	
PHOX2B	Ch	cg05323533	4	8/20	40.00%	
PITX1	Ch	cg00396667	5	6/20	30.00%	
PITX2	Ch	cg08979895	4	6/20	30.00%	McRonald.F et al 2009
PM20D1	Ch	cg11965913	1	8/20	40.00%	
PPP2R2A	Ch	cg08201604	8	7/20	35.00%	
PRDM13	Ch	cg01815538	6	6/20	30.00%	
PRDM16	Ch	cg03254465	1	8/20	40.00%	
PRDM16	Ch	cg09990962	1	10/20	50.00%	
PRDM16	Ch	cg14200569	1	10/20	50.00%	
PRKCDBP	Ch	cg26678920	11	7/20	35.00%	
PROCA1	Ch	cg02685896	17	11/20	55.00%	
PROCA1	Ch	cg14676825	17	10/20	50.00%	
PRR5	Ch	cg24019054	22	7/20	35.00%	

Gene	Cancer	TargetID	CHR	Prevelance		Frequently Methylated in other Kidney Cancers
PTHLH	Ch	cg10253371	12	8/20	40.00%	
PTPRD	Ch	cg09371281	9	6/20	30.00%	
PTPRD	Ch	cg14458619	9	7/20	35.00%	
RAB37	Ch	cg16347279	17	6/20	30.00%	
RALYL	Ch	cg22403811	8	6/20	30.00%	
RASA3	Ch	cg14696311	13	11/20	55.00%	
RAX	Ch	cg03671802	18	6/20	30.00%	
RBM20	Ch	cg08692733	10	6/20	30.00%	
RRBP1	Ch	cg26447697	20	6/20	30.00%	
SALL1	Ch	cg08304084	16	12/20	60.00%	
SCMH1	Ch	cg00826203	1	6/20	30.00%	
SDK1	Ch	cg16956686	7	9/20	45.00%	
SDK1	Ch	cg24796546	7	8/20	40.00%	
SDK2	Ch	cg06061257	17	7/20	35.00%	
SEMA4A	Ch	cg05047401	1	6/20	30.00%	
SH3PXD2A	Ch	cg18735015	10	6/20	30.00%	
SH3RF3	Ch	cg18145759	2	7/20	35.00%	
SHANK2	Ch	cg26591930	11	7/20	35.00%	
SIAH3	Ch	cg26667946	13	6/20	30.00%	
SIM1	Ch	cg08074534	6	6/20	30.00%	
SIX2	Ch	cg24887265	2	6/20	30.00%	
SKAP2	Ch	cg03730533	7	7/20	35.00%	
SLC16A12	Ch	cg03616221	10	7/20	35.00%	
SLC2A14	Ch	cg06645921	12	9/20	45.00%	
SLC2A3	Ch	cg20972214	12	8/20	40.00%	
SLC2A9	Ch	cg25117600	4	10/20	50.00%	
SLFN12L	Ch	cg22763680	17	6/20	30.00%	
SLITRK1	Ch	cg16727923	13	7/20	35.00%	
SOX1	Ch	cg16705627	13	6/20	30.00%	
SOX2OT	Ch	cg24513480	3	8/20	40.00%	
SPG20	Ch	cg10558887	13	7/20	35.00%	
SRCIN1	Ch	cg19409060	17	7/20	35.00%	
TAF13	Ch	cg20569839	1	6/20	30.00%	
TBCD	Ch	cg00960700	17	6/20	30.00%	
TCTEX1D1	Ch	cg09769134	1	7/20	35.00%	
TET1	Ch	cg02952701	10	8/20	40.00%	
TET1	Ch	cg22876739	10	7/20	35.00%	
TFAP2B	Ch	cg00908833	6	6/20	30.00%	
TFAP2B	Ch	cg03159396	6	12/20	60.00%	
TFAP2B	Ch	cg05437823	6	8/20	40.00%	
TFAP2B	Ch	cg08178940	6	14/20	70.00%	
TFAP2B	Ch	cg08857063	6	8/20	40.00%	

Gene	Cancer	TargetID	CHR	Prevelance		Frequently Methylated in other Kidney Cancers
TFAP2B	Ch	cg22282405	6	9/20	45.00%	
THBS4	Ch	cg19025234	5	6/20	30.00%	
TLX3	Ch	cg25942450	5	7/20	35.00%	Ricketts et al 2012
TMC2	Ch	cg07341624	20	6/20	30.00%	
TMEM171	Ch	cg10795659	5	9/20	45.00%	
TMEM37	Ch	cg09474442	2	6/20	30.00%	
TMEM37	Ch	cg09715353	2	9/20	45.00%	
TOX2	Ch	cg10900455	20	6/20	30.00%	
TRIL	Ch	cg02624770	7	6/20	30.00%	
TRIM10	Ch	cg16912957	6	8/20	40.00%	
TRIM10	Ch	cg18420781	6	6/20	30.00%	
TRIM46	Ch	cg15497761	1	6/20	30.00%	
TRIM59	Ch	cg15618978	3	6/20	30.00%	
TUBGCP6	Ch	cg19851816	22	7/20	35.00%	
TULP1	Ch	cg24281267	6	10/20	50.00%	
TXNRD1	Ch	cg09884423	12	6/20	30.00%	
VPS37B	Ch	cg16953816	12	18/20	90.00%	
VSX1	Ch	cg02898293	20	6/20	30.00%	
ZC3H12D	Ch	cg09313931	6	7/20	35.00%	
ZC3H12D	Ch	cg10308253	6	7/20	35.00%	
ZC3H12D	Ch	cg15559674	6	11/20	55.00%	
ZEB2	Ch	cg09596336	2	6/20	30.00%	
ZFHX3	Ch	cg08512490	16	8/20	40.00%	
ZFHX3	Ch	cg16630989	16	7/20	35.00%	
ZIC1	Ch	cg05095591	3	6/20	30.00%	
ZIC2	Ch	cg24742746	13	7/20	35.00%	
ZIC4	Ch	cg01137401	3	7/20	35.00%	
ZIC4	Ch	cg04556126	3	6/20	30.00%	
ZNF177	Ch	cg09578475	19	6/20	30.00%	Ricketts et al 2012
ZNF83	Ch	cg17132967	19	6/20	30.00%	

Table 7-7: Hypermethylated gene in chromophobe RCC, β value > 0.5 in > 30% samples.

Gene list of all probes methylated in >30% oncocytoma Samples

Gene	Cancer	TargetID	CHR	Prevelance		Frequently Methylated in other Kidney Cancers
ACBD5	Onco	cg14240646	10	7/21	33.33%	
ACN9	Onco	cg10548492	7	7/21	33.33%	
ACP5	Onco	cg04566159	19	7/21	33.33%	
ADAMTS14	Onco	cg04838832	10	7/21	33.33%	
AFF1	Onco	cg17455261	4	9/21	42.86%	
AGAP1	Onco	cg17945789	2	13/21	61.90%	
ALCAM	Onco	cg05645404	3	12/21	57.14%	
ARHGAP10	Onco	cg11713658	4	12/21	57.14%	
ASB4	Onco	cg02805025	7	11/21	52.38%	
ATP6V1B1	Onco	cg04751811	2	14/21	66.67%	
ATP6V1B1	Onco	cg06288697	2	15/21	71.43%	
C14orf64	Onco	cg16278496	14	7/21	33.33%	
C2orf43	Onco	cg14757738	2	9/21	42.86%	
CAPN2	Onco	cg06756211	1	12/21	57.14%	
CCND3	Onco	cg04111789	6	9/21	42.86%	
CMTM3	Onco	cg26560414	16	8/21	38.10%	
COL4A1	Onco	cg23008352	13	11/21	52.38%	
CUX1	Onco	cg02856420	7	7/21	33.33%	
CXCL14	Onco	cg04002608	5	9/21	42.86%	
ELTD1	Onco	cg15084543	1	11/21	52.38%	
EPS8L1	Onco	cg08166750	19	8/21	38.10%	
EPS8L1	Onco	cg12020396	19	12/21	57.14%	
EPS8L1	Onco	cg18320766	19	15/21	71.43%	
EPS8L2	Onco	cg08327690	11	10/21	47.62%	
GAD1	Onco	cg04105250	2	9/21	42.86%	
GIPC2	Onco	cg01074657	1	7/21	33.33%	
GRIK2	Onco	cg21635870	6	10/21	47.62%	
HOXA2	Onco	cg02225599	7	12/21	57.14%	
HOXA3	Onco	cg21134232	7	9/21	42.86%	
HOXC10	Onco	cg20402783	12	8/21	38.10%	
KCNQ4	Onco	cg19689322	1	9/21	42.86%	
MAML2	Onco	cg24088496	11	15/21	71.43%	
MFHAS1	Onco	cg27280535	8	7/21	33.33%	
MIR196A1	Onco	cg26608174	17	10/21	47.62%	
NFE2L3	Onco	cg10536999	7	10/21	47.62%	
NRG2	Onco	cg02009088	5	14/21	66.67%	
NRG2	Onco	cg05652757	5	11/21	52.38%	
NRG2	Onco	cg10468961	5	9/21	42.86%	
NRG2	Onco	cg15992535	5	16/21	76.19%	
NRG2	Onco	cg22060611	5	8/21	38.10%	
PACS2	Onco	cg22976533	14	8/21	38.10%	
PARD3	Onco	cg15292765	10	17/21	80.95%	
PARD3	Onco	cg16895672	10	7/21	33.33%	

Gene	Cancer	TargetID	CHR	Prevelance		Frequently Methylated in other Kidney Cancers
PCDHB17	Onco	cg21777188	5	9/21	42.86%	
PCDHGA2	Onco	cg18781988	5	7/21	33.33%	
PCDHGA4	Onco	cg21627409	5	9/21	42.86%	
PCSK6	Onco	cg07324702	15	9/21	42.86%	
PDE4D	Onco	cg03323696	5	8/21	38.10%	
PHACTR1	Onco	cg13246235	6	7/21	33.33%	
PRDM16	Onco	cg03254465	1	8/21	38.10%	
PRDM16	Onco	cg09990962	1	12/21	57.14%	
PRDM16	Onco	cg14200569	1	10/21	47.62%	
PRDM16	Onco	cg24939838	1	7/21	33.33%	
PROCA1	Onco	cg02685896	17	8/21	38.10%	
PROCA1	Onco	cg14676825	17	10/21	47.62%	
PYY	Onco	cg16789592	17	8/21	38.10%	
RPH3A	Onco	cg01573562	12	9/21	42.86%	
SALL1	Onco	cg08304084	16	16/21	76.19%	
SCMH1	Onco	cg00826203	1	10/21	47.62%	
SDS	Onco	cg14120215	12	8/21	38.10%	
SHANK2	Onco	cg26591930	11	12/21	57.14%	
SIM1	Onco	cg08074534	6	8/21	38.10%	
SKAP2	Onco	cg03730533	7	16/21	76.19%	
SLC16A12	Onco	cg03616221	10	11/21	52.38%	
SLC2A9	Onco	cg25117600	4	13/21	61.90%	
SPTBN4	Onco	cg07063351	19	9/21	42.86%	
SPTBN4	Onco	cg08871964	19	7/21	33.33%	
SRCIN1	Onco	cg15809959	17	8/21	38.10%	
TCTEX1D1	Onco	cg09769134	1	7/21	33.33%	
TCTEX1D4	Onco	cg03434029	1	8/21	38.10%	
TCTEX1D4	Onco	cg11448683	1	8/21	38.10%	
TFAP2A	Onco	cg00689580	6	12/21	57.14%	
TFAP2A	Onco	cg10129408	6	13/21	61.90%	
THBS4	Onco	cg19025234	5	10/21	47.62%	
TMEM101	Onco	cg16182148	17	7/21	33.33%	
TMEM37	Onco	cg09715353	2	7/21	33.33%	
TRIM10	Onco	cg16912957	6	10/21	47.62%	
TRIM10	Onco	cg18420781	6	7/21	33.33%	
TRPC4AP	Onco	cg01154966	20	7/21	33.33%	
VAT1L	Onco	cg04591018	16	9/21	42.86%	
VPS37B	Onco	cg16953816	12	17/21	80.95%	
WFDC10A	Onco	cg15107861	20	7/21	33.33%	
ZBED3	Onco	cg03163459	5	10/21	47.62%	
ZC3H12D	Onco	cg09313931	6	8/21	38.10%	
ZC3H12D	Onco	cg15559674	6	9/21	42.86%	
ZIC4	Onco	cg23189410	3	7/21	33.33%	

Table 7-8: Hypermethylated gene in renal oncocytoma, β value > 0.5 in > 30% samples.

Chapter Eight: **References**

Adibi, M., D, M., Thomas, A.Z., D, M., Borregales, L.D., D, M., Merrill, M.M., O, D., Slack, R.S., S, M., Chen, H., S, M., Sircar, K., D, M., Murugan, P., D, M., Tamboli, P., D, M., Jonasch, E., D, M., Tannir, N.M., P, F.A.C., Matin, S.F., S, F.A.C., Wood, C.G., S, F.A.C., Karam, J.A., S, F.A.C., 2015. Percentage of sarcomatoid component as a prognostic indicator for survival in renal cell carcinoma with sarcomatoid dedifferentiation. Urol. Oncol. Semin. Orig. Investig. 1–7. doi:10.1016/j.urolonc.2015.04.011

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–9. doi:10.1038/nmeth0410-248

Agathanggelou, A., Cooper, W.N., Latif, F., 2005. Role of the Ras-Association Domain Family 1 Tumor Suppressor Gene in Human Cancers. Cancer Res. 65, 3497–3508. doi:10.1158/0008-5472.CAN-04-4088

Aghaallaei, N., Bajoghli, B., Czerny, T., 2007. Distinct roles of Fgf8, Foxi1, Dlx3b and Pax8/2 during otic vesicle induction and maintenance in medaka. Dev. Biol. 307, 408–20. doi:10.1016/j.ydbio.2007.04.022

Ahn, E.Y., Kim, J.S., Kim, G.J., Park, Y.N., 2013. RASSF1A-mediated regulation of AREG via the Hippo pathway in hepatocellular carcinoma. Mol. cancer Res. 11, 748–58. doi:10.1158/1541-7786.MCR-12-0665

Akbani, R., Ng, P.K.S., Werner, H.M.J., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J.-Y., Yoshihara, K., Li, J., Ling, S., Seviour, E.G., Ram, P.T., Minna, J.D., Diao, L., Tong, P., Heymach, J. V, Hill, S.M., Dondelinger, F., Städler, N., Byers, L. a, Meric-Bernstam, F., Weinstein, J.N., Broom, B.M., Verhaak, R.G.W., Liang, H., Mukherjee, S., Lu, Y., Mills, G.B., 2014. A pan-cancer proteomic perspective on The Cancer Genome Atlas. Nat. Commun. 5, 3887. doi:10.1038/ncomms4887

Alazami, a. M., Alshammari, M.J., Baig, M., Salih, M. a., Hassan, H.H., Alkuraya, F.S., 2014. NPHP4 mutation is linked to cerebello-oculo-renal syndrome and male infertility. Clin. Genet. 85, 371–375. doi:10.1111/cge.12160

Albiges, L., Guegan, J., Le Formal, A., Verkarre, V., Rioux-Leclercq, N., Sibony, M., Bernhard, J.-C., Camparo, P., Merabet, Z., Molinie, V., Allory, Y., Orear, C., Couve, S., Gad, S., Patard, J.-J., Escudier, B., 2014. MET is a potential target across all Papillary Renal Cell Carcinomas. Result from a large molecular study of pRCC with CGHa and matching Gene Expression array. Clin. Cancer Res. 3411–3422. doi:10.1158/1078-0432.CCR-13-2173

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., et al. 2013. Signatures of mutational processes in human cancer. Nature 500, 415–421. doi:10.1038/nature12477

Alholle, A., Brini, A.T., Bauer, J., Gharanei, S., Niada, S., Slater, A., Gentle, D., Maher, E.R., Jeys, L., Grimer, R., Sumathi, V.P., Latif, F., 2015. Genome-wide DNA methylation profiling of recurrent and non-recurrent chordomas. Epigenetics 10, 213–20. doi:10.1080/15592294.2015.1006497

Antoniou, A.C., Pharoah, P.P.D., Smith, P., Easton, D.F., 2004. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br. J. Cancer 91, 1580–90. doi:10.1038/sj.bjc.6602175

Arai, E., Chiku, S., Mori, T., Gotoh, M., Nakagawa, T., Fujimoto, H., Kanai, Y., 2012. Single-CpG-resolution methylome analysis identifies clinicopathologically aggressive CpG island methylator phenotype clear cell renal cell carcinomas. Carcinogenesis 33, 1487–93. doi:10.1093/carcin/bgs177

Arai, E., Wakai-Ushijima, S., Fujimoto, H., Hosoda, F., Shibata, T., Kondo, T., Yokoi, S., Imoto, I., Inazawa, J., Hirohashi, S., Kanai, Y., 2011. Genome-wide DNA methylation profiles in renal tumors of various histological subtypes and non-tumorous renal tissues. Pathobiology 78, 1–

9. doi:10.1159/000322072

Argos, M., 2015. Arsenic Exposure and Epigenetic Alterations: Recent Findings Based on the Illumina 450K DNA Methylation Array. Curr. Environ. Heal. reports 2, 137–44. doi:10.1007/s40572-015-0052-1

Avissar-Whiting, M., Koestler, D.C., Houseman, E.A., Christensen, B.C., Kelsey, K.T., Marsit, C.J., 2011. Polycomb group genes are targets of aberrant DNA methylation in renal cell carcinoma. Epigenetics 6, 703–9.

Avvakumov, N., Côté, J., 2007. The MYST family of histone acetyltransferases and their intimate links to cancer. Oncogene 26, 5395–407. doi:10.1038/sj.onc.1210608

Azad, N., Zahnow, C.A., Rudin, C.M., Baylin, S.B., 2013. The future of epigenetic therapy in solid tumours--lessons from the past. Nat. Rev. Clin. Oncol. 10, 256–66. doi:10.1038/nrclinonc.2013.42

Baert, A.L., Sartor, K., 2006. Imaging of Kidney Cancer. Springer Science & Business Media.

Bao, Y., Hata, Y., Ikeda, M., Withanage, K., 2011. Mammalian Hippo pathway: from development to cancer and beyond. J. Biochem. 149, 361–79. doi:10.1093/jb/mvr021

Bayés, M., Heath, S., Gut, I.G., 2011. Applications of Second Generation Sequencing Technologies in Complex Disorders. Springer Berlin Heidelberg, pp. 321–343. doi:10.1007/7854_2011_196

Baylin, S.B., Jones, P.A., 2011. A decade of exploring the cancer epigenome — biological and translational implications.

Bell, D.W., 2010. Our changing view of the genomic landscape of cancer. J. Pathol. 220, 231–43. doi:10.1002/path.2645

Benn, M., Tybjærg-Hansen, A., Smith, G.D., Nordestgaard, B.G., 2016. High body mass index and cancer risk-a Mendelian randomisation study. Eur. J. Epidemiol. doi:10.1007/s10654-016-0147-5

Berg, M., Agesen, T.H., Thiis-Evensen, E., Merok, M.A., Teixeira, M.R., Vatn, M.H., Nesbakken, A., Skotheim, R.I., Lothe, R.A., 2010. Distinct high resolution genome profiles of early onset and late onset colorectal cancer integrated with gene expression data identify candidate susceptibility loci. Mol. Cancer 9, 100. doi:10.1186/1476-4598-9-100

Bethge, N., Lothe, R.A., Honne, H., Andresen, K., Trøen, G., Eknæs, M., Liestøl, K., Holte, H., Delabie, J., Smeland, E.B., Lind, G.E., 2014. Colorectal cancer DNA methylation marker panel validated with high performance in Non-Hodgkin lymphoma. Epigenetics 9, 428–36. doi:10.4161/epi.27554

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., Fan, J.B., Shen, R., 2011. High density DNA methylation array with single CpG site resolution. Genomics 98, 288–295. doi:10.1016/j.ygeno.2011.07.007

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E.W., Wu, B., Doucet, D., Thomas, N.J., Wang, Y., Vollmer, E., Goldmann, T., Seifart, C., Jiang, W., Barker, D.L., Chee, M.S., Floros, J., Fan, J.-B., 2006. High-throughput DNA methylation profiling using universal bead arrays. Genome Res. 16, 383–93. doi:10.1101/gr.4410706

Blomqvist, S.R., Vidarsson, H., Fitzgerald, S., Johansson, B.R., Ollerstam, A., Brown, R., Persson, A.E.G., Bergström G, G. öran, Enerbäck, S., 2004. Distal renal tubular acidosis in mice that lack the forkhead transcription factor Foxi1. J. Clin. Invest. 113, 1560–70. doi:10.1172/JCI20665

Bodmer, D., van den Hurk, W., van Groningen, J.J.M., Eleveld, M.J., Martens, G.J.M., Weterman, M. a J., van Kessel, A.G., 2002. Understanding familial and non-familial renal cell cancer. Hum. Mol. Genet. 11, 2489–2498.

Boichard, A., Croux, L., Al Ghuzlan, A., Broutin, S., Dupuy, C., Leboulleux, S., Schlumberger, M., Bidart, J.M., Lacroix, L., 2012. Somatic RAS mutations occur in a large proportion of sporadic

RET-negative medullary thyroid carcinomas and extend to a previously unidentified exon. J. Clin. Endocrinol. Metab. 97, E2031–5. doi:10.1210/jc.2012-2092

Borah, S., Xi, L., Zaug, A.J., Powell, N.M., Dancik, G.M., Cohen, S.B., Costello, J.C., Theodorescu, D., Cech, T.R., 2015. TERT promoter mutations and telomerase reactivation in urothelial cancer. Science (80-.). 347, 1006–10. doi:10.1126/science.1260200

Borowicz, S., Van Scoyk, M., Avasarala, S., Karuppusamy Rathinam, M.K., Tauler, J., Bikkavilli, R.K., Winn, R. a., 2014. The Soft Agar Colony Formation Assay. J. Vis. Exp. 1–6. doi:10.3791/51998

Brena, R.M., Plass, C., 2009. DNA Methylation. doi:10.1007/978-1-59745-522-0

Brinker, D.A., Amin, M.B., de Peralta-Venturina, M., Reuter, V., Chan, D.Y., Epstein, J.I., 2000. Extensively necrotic cystic renal cell carcinoma: a clinicopathologic study with comparison to other cystic and necrotic renal cancers. Am. J. Surg. Pathol. 24, 988–95.

Brock, M. V, Herman, J.G., Baylin, S.B., 2007. Cancer as a manifestation of aberrant chromatin structure. Cancer J 13, 3–8. doi:10.1097/PPO.0b013e31803c5415

Brugarolas, J., 2014. Molecular Genetics of Clear-Cell Renal Cell Carcinoma. J. Clin. Oncol. 1–10. doi:10.1200/JCO.2012.45.2003

Buhmeida, A., Merdad, A., Al-Maghrabi, J., El-Maghrabi, J., Al-Thobaiti, F., Ata, M., Bugis, A., Syrjänen, K., Abuzenadah, A., Chaudhary, A., Gari, M., Al-Qahtani, M., Dallol, A., 2011. RASSF1A methylation is predictive of poor prognosis in female breast cancer in a background of overall low methylation frequency. Anticancer Res. 31, 2975–81.

Burrell, R.A., McGranahan, N., Bartek, J., Swanton, C., 2013. The causes and consequences of genetic heterogeneity in cancer evolution. Nature 501, 338–45. doi:10.1038/nature12625

Busche, S., Shao, X., Caron, M., et. al 2015. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. Genome Biol. 16, 290. doi:10.1186/s13059-015-0856-1

Bylow, K.A., Atkins, M.B., Posadas, E.M., Stadler, W.M., McDermott, D.F., 2009. Phase II trial of carboplatin and paclitaxel in papillary renal cell carcinoma. Clin. Genitourin. Cancer 7, 39–42. doi:10.3816/CGC.2009.n.007

Cairns, P., 2010. Renal cell carcinoma. Cancer Biomark. 9, 461–73. doi:10.3233/CBM-2011-0176

Calnan, D.R., Brunet, A., 2008. The FoxO code. Oncogene 27, 2276–2288. doi:10.1038/onc.2008.21

Cao, R., Zhang, Y., 2004. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol. Cell 15, 57–67. doi:10.1016/j.molcel.2004.06.020

Carr, I.M., Valleley, E.M.A., Cordery, S.F., Markham, A.F., Bonthron, D.T., 2007. Sequence analysis and editing for bisulphite genomic sequencing projects. Nucleic Acids Res. 35, e79. doi:10.1093/nar/gkm330

Cavallo, F., De Giovanni, C., Nanni, P., Forni, G., Lollini, P.-L., 2011. 2011: the Immune Hallmarks of Cancer. Cell 60, 319–326. doi:10.1016/S0092-8674(00)81683-9

Ceccacci, E., Minucci, S., 2016. Inhibition of histone deacetylases in cancer therapy: lessons from leukaemia. Br. J. Cancer 114, 605–611. doi:10.1038/bjc.2016.36

Cedar, H., Bergman, Y., 2009. Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. 10, 295–304. doi:10.1038/nrg2540

Cetin, E., Cengiz, B., Gunduz, E., Gunduz, M., Nagatsuka, H., Bekir-Beder, L., Fukushima, K., Pehlivan, D., N, M.O., Nishizaki, K., Shimizu, K., Nagai, N., 2008. Deletion mapping of chromosome 4q22-35 and identification of four frequently deleted regions in head and neck cancers. Neoplasma 55, 299–304.

Chan, J.J., Flatters, D., Rodrigues-Lima, F., Yan, J., Thalassinos, K., Katan, M., 2013. Comparative analysis of interactions of RASSF1-10. Adv. Biol. Regul. 53, 190–201. doi:10.1016/j.jbior.2012.12.001

Chen, J., Lui, W.O., Vos, M.D., Clark, G.J., Takahashi, M., Schoumans, J., Khoo, S.K., Petillo, D., Lavery, T., Sugimura, J., Astuti, D., Zhang, C., Kagawa, S., Maher, E.R., Larsson, C., Alberts, A.S., Kanayama, H.O., Teh, B.T., 2003. The t(1;3) breakpoint-spanning genes LSAMP and NORE1 are involved in clear cell renal cell carcinomas. Cancer Cell 4, 405–413. doi:10.1016/S1535-6108(03)00269-1

Chen, K., He, J., Wang, D., Cao, J., Li, M., Zhao, X., Sheng, X., Li, W., Liu, W., 2014. Methylation□associated inactivation of LATS1 and its effect on demethylation or overexpression on YAP and cell biological function in human renal cell carcinoma. Int. J. Oncol. 45, 2511–2521.

Chen, Y., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., Weksberg, R., 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8, 203–9. doi:10.4161/epi.23470

Chen, Y., Liu, W., Naumovski, L., Neve, R.L., 2003. ASPP2 inhibits APP-BP1-mediated NEDD8 conjugation to cullin-1 and decreases APP-BP1-induced cell proliferation and neuronal apoptosis. J. Neurochem. 85, 801–809. doi:10.1046/j.1471-4159.2003.01727.x

Chmielecki, J., Meyerson, M., 2014. DNA Sequencing of Cancer: What Have We Learned? Annu. Rev. Med. 65, 63–79. doi:10.1146/annurev-med-060712-200152

Chuang, J.C., Jones, P.A., 2007. Epigenetics and microRNAs. Pediatr. Res. 61, 24R–29R. doi:10.1203/pdr.0b013e3180457684

Clifford, S.C., Maher, E.R., 2001. Von Hippel-Lindau disease: clinical and molecular perspectives. Adv. Cancer Res. 82, 85–105.

Cline, M.S., Craft, B., Swatloski, T., Goldman, M., Ma, S., Haussler, D., Zhu, J., 2013. Exploring TCGA Pan-Cancer Data at the UCSC Cancer Genomics Browser. Sci. Rep. 3, 2652. doi:10.1038/srep02652

Coleman, W.B., Rivenbark, A.G., 2006. Quantitative DNA methylation analysis: the promise of high-throughput epigenomic diagnostic testing in human neoplastic disease. J. Mol. Diagn. 8, 152–6. doi:10.2353/jmoldx.2006.060026

Conner, J.R., Hirsch, M.S., Jo, V.Y., 2015. HNF1β and S100A1 are useful biomarkers for distinguishing renal oncocytoma and chromophobe renal cell carcinoma in FNA and core needle biopsies. Cancer Cytopathol. 123, 298–305. doi:10.1002/cncy.21530

Cooper, G.M., Shendure, J., 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat. Rev. Genet. 12, 628–40. doi:10.1038/nrg3046

Cooper, W.N., Hesson, L.B., Matallanas, D., Dallol, A., von Kriegsheim, A., Ward, R., Kolch, W., Latif, F., 2009. RASSF2 associates with and stabilizes the proapoptotic kinase MST2. Oncogene 28, 2988–2998. doi:10.1038/onc.2009.152

Costa, V.L., Henrique, R., Danielsen, S. a., Eknaes, M., Patrício, P., Morais, A., Oliveira, J., Lothe, R. a., Teixeira, M.R., Lind, G.E., Jerónimo, C., 2011. TCF21 and PCDH17 methylation: An innovative panel of biomarkers for a simultaneous detection of urological cancers. Epigenetics 6, 1120–1130. doi:10.4161/epi.6.9.16376

Costa, V.L., Henrique, R., Ribeiro, F.R., Pinto, M., Oliveira, J., Lobo, F., Teixeira, M.R., Jerónimo, C., 2007. Quantitative promoter methylation analysis of multiple cancer-related genes in renal cell tumors. BMC Cancer 7, 133. doi:10.1186/1471-2407-7-133

Costello, J.F., Frühwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomäki, P., Lang, J.C., Schuller, D.E., Yu, L., Bloomfield, C.D., Caligiuri, M.A., Yates, A., Nishikawa, R., Su Huang, H., Petrelli, N.J., Zhang, X., O'Dorisio, M.S., Held, W.A.,

Cavenee, W.K., Plass, C., 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat. Genet. 24, 132–8. doi:10.1038/72785

Courthod, G., Tucci, M., Di Maio, M., Scagliotti, G. V., 2015. Papillary renal cell carcinoma: A review of the current therapeutic landscape. Crit. Rev. Oncol. Hematol. doi:10.1016/j.critrevonc.2015.05.008

Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., Teague, J., Andrews, J., Barthorpe, S., Beare, D., Buck, G., Campbell, P.J., Forbes, S., Jia, M., Jones, D., Knott, H., Kok, C.Y., Lau, K.W., Leroy, C., Lin, M.-L., McBride, D.J., Maddison, M., Maguire, S., McLay, K., Menzies, A., Mironenko, T., Mulderrig, L., Mudie, L., O'Meara, S., Pleasance, E., Rajasingham, A., Shepherd, R., Smith, R., Stebbings, L., Stephens, P., Tang, G., Tarpey, P.S., Turrell, K., Dykema, K.J., Khoo, S.K., Petillo, D., Wondergem, B., Anema, J., Kahnoski, R.J., Teh, B.T., Stratton, M.R., Futreal, P.A., 2010. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 463, 360–363. doi:10.1038/nature08672

Dallol, A., Al-Ali, W., Al-Shaibani, A., Fahd, A.-M., Al-Mulla, F., 2011. Analysis of DNA methylation in FFPE tissues using the MethyLight technology. Methods Mol. Biol. 724, 191–204. doi:10.1007/978-1-61779-055-3_13

Dammann, R., Li, C., Yoon, J.H., Chin, P.L., Bates, S., Pfeifer, G.P., 2000. Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat. Genet. 25, 315–319. doi:10.1038/77083

Davidović, R.S., Božović, A.M., Mandušić, V.L., Krajnović, M.M., 2014. Methylation-specific PCR: four steps in primer design. Open Life Sci. 9, 1127–1139. doi:10.2478/s11535-014-0324-z

Davis, C.F.F., Ricketts, C.J., Wang, M., Yang, L., Cherniack, A.D.D., Shen, H., Buhay, C., Kang, H., Kim, S.C.C., Fahey, C.C.C., Hacker, K.E.E., Bhanot, G., Gordenin, D.A.A., Chu, A., Gunaratne, P.H.H., Biehl, M., Seth, S., Kaipparettu, B.A.A., Bristow, C.A.A., Donehower, L.A.A., Wallen, E.M.M., Smith, A.B.B., Tickoo, S.K.K., Tamboli, P., Reuter, V., Schmidt, L.S.S., Hsieh, J.J.J., Choueiri, T.K.K., Hakimi, a. A.A., Chin, L., Meyerson, M., Kucherlapati, R., Park, W.-Y.Y., Robertson, a. G., Laird, P.W., Henske, E.P., Kwiatkowski, D.J., Park, P.J., Morgan, M., Shuch, B., Muzny, D., Wheeler, D. a., Linehan, W.M., Gibbs, R. a., Rathmell, W.K., Creighton, C.J., 2014. The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer Cell 26, 319–30. doi:10.1016/j.ccr.2014.07.014

de Hoon, M.J.L., Imoto, S., Nolan, J., Miyano, S., 2004. Open source clustering software. Bioinformatics 20, 1453–4. doi:10.1093/bioinformatics/bth078

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., Fuks, F., 2011. Evaluation of the Infinium Methylation 450K technology. Epigenomics 3, 771–84. doi:10.2217/epi.11.105

Delahunt, B., Cheville, J.C., Martignoni, G., Humphrey, P.A., Magi-Galluzzi, C., McKenney, J., Egevad, L., Algaba, F., Moch, H., Grignon, D.J., Montironi, R., Srigley, J.R., 2013. The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. Am. J. Surg. Pathol. 37, 1490–504. doi:10.1097/PAS.0b013e318299f0fb

Dietel, M., Jöhrens, K., Laffert, M., Hummel, M., Bläker, H., Müller, B.M., Lehmann, A., Denkert, C., Heppner, F.L., Koch, A., Sers, C., Anagnostopoulos, I., 2015. A 2015 update on predictive molecular pathology and its role in targeted cancer therapy: a review focussing on clinical relevance. Cancer Gene Ther. 20, 211–21. doi:10.1038/cgt.2013.13

Dimauro, T., David, G., 2010. Ras-induced senescence and its physiological relevance in cancer. Curr. Cancer Drug Targets 10, 869–76.

Dittfeld, C., Richter, A.M., Steinmann, K., Klagge-Ulonska, A., Dammann, R.H., 2012. The SARAH Domain of RASSF1A and Its Tumor Suppressor Function. Mol. Biol. Int. 2012, 196715. doi:10.1155/2012/196715

Dong, X., He, H., Zhang, W., Yu, D., Wang, X., Chen, Y., 2015. Combination of serum RASSF1A methylation and AFP is a promising non-invasive biomarker for HCC patient with chronic HBV infection. Diagn. Pathol. 10, 133. doi:10.1186/s13000-015-0317-x

Donninger, H., Clark, J. a., Monaghan, M.K., Schmidt, M.L., Vos, M., Clark, G.J., 2014. Cell Cycle Restriction Is More Important Than Apoptosis Induction for RASSF1A Protein Tumor Suppression. J. Biol. Chem. 289, 31287–31295. doi:10.1074/jbc.M114.609537

Donninger, H., Vos, M.D., Clark, G.J., 2007. The RASSF1A tumor suppressor. J. Cell Sci. 120, 3163–3172. doi:10.1242/jcs.010389

Du, P., Kibbe, W.A., Lin, S.M., 2008. lumi: a pipeline for processing Illumina microarray. Bioinformatics 24, 1547–8. doi:10.1093/bioinformatics/btn224

Dulaimi, E., Caceres, I.I. De, Uzzo, R.G., Al-saleem, T., Greenberg, R.E., Polascik, T.J., Babb, J.S., Grizzle, W.E., Cairns, P., 2004. Promoter Hypermethylation Profile of Kidney Cancer Promoter Hypermethylation Profile of Kidney Cancer. Clin Cancer Res 10, 3972–3979.

Dunlop, E.A., Seifan, S., Claessens, T., Behrends, C., Kamps, M.A., Rozycka, E., Kemp, A.J., Nookala, R.K., Blenis, J., Coull, B.J., Murray, J.T., van Steensel, M.A., Wilkinson, S., Tee, A.R., 2014. FLCN, a novel autophagy component, interacts with GABARAP and is regulated by ULK1 phosphorylation. Autophagy 10, 1749–60. doi:10.4161/auto.29640

Durinck, S., Stawiski, E.W., Pavía-Jiménez, A., Modrusan, Z., Kapur, P., Jaiswal, B.S., Zhang, N., Toffessi-Tcheuyap, V., Nguyen, T.T., Pahuja, K.B., Chen, Y.-J., Saleem, S., Chaudhuri, S., Heldens, S., Jackson, M., Peña-Llopis, S., Guillory, J., Toy, K., Ha, C., Harris, C.J., Holloman, E., Hill, H.M., Stinson, J., Rivers, C.S., Janakiraman, V., Wang, W., Kinch, L.N., Grishin, N. V, Haverty, P.M., Chow, B., Gehring, J.S., Reeder, J., Pau, G., Wu, T.D., Margulis, V., Lotan, Y., Sagalowsky, A., Pedrosa, I., de Sauvage, F.J., Brugarolas, J., Seshagiri, S., 2015. Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes. Nat. Genet. 47, 13–21. doi:10.1038/ng.3146

Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P. V, Laird, P.W., 2000. MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. 28, E32.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T. V, Davies, R., Down, T.A., Haefliger, C., Horton, R., Howe, K., Jackson, D.K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K., Beck, S., 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. 38, 1378–85. doi:10.1038/ng1909

Egner, J.R., 2010. AJCC Cancer Staging Manual. JAMA J. Am. Med. Assoc. doi:10.1001/jama.2010.1525

Ehrlich, M., 2009. DNA hypomethylation in cancer cells. Epigenomics 1, 239–59. doi:10.2217/epi.09.33

Ehrlich, M., 2002. DNA methylation in cancer: too much, but also too little. Oncogene 21, 5400–5413. doi:10.1038/sj.onc.1205651

El-Dahr, S.S., Aboudehen, K., Saifudeen, Z., 2008. Transcriptional control of terminal nephron differentiation. Am. J. Physiol. Renal Physiol. 294, F1273–8. doi:10.1152/ajprenal.00562.2007

Ellinger, J., Holl, D., Nuhn, P., Kahl, P., Haseke, N., Staehler, M., Siegert, S., Hauser, S., Stief, C.G., Müller, S.C., Bastian, P.J., 2011a. DNA hypermethylation in papillary renal cell carcinoma. BJU Int. 107, 664–9. doi:10.1111/j.1464-410X.2010.09468.x

Ellinger, J., Holl, D., Nuhn, P., Kahl, P., Haseke, N., Staehler, M., Siegert, S., Hauser, S., Stief, C.G., Müller, S.C., Bastian, P.J., 2011b. DNA hypermethylation in papillary renal cell carcinoma. BJU Int. 107, 664–669. doi:10.1111/j.1464-410X.2010.09468.x

Ellinger, J., Kahl, P., Mertens, C., Rogenhofer, S., Hauser, S., Hartmann, W., Bastian, P.J., Büttner,

R., Müller, S.C., von Ruecker, A., 2010. Prognostic relevance of global histone H3 lysine 4 (H3K4) methylation in renal cell carcinoma. Int. J. Cancer 127, 2360–2366. doi:10.1002/ijc.25250

Escudier, B., 2012. Emerging immunotherapies for renal cell carcinoma. Ann. Oncol. 23 Suppl 8, viii35–40. doi:10.1093/annonc/mds261

Escudier, B., Eisen, T., Porta, C., Patard, J.J., Khoo, V., Algaba, F., Mulders, P., Kataja, V., 2012. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. 23 Suppl 7, vii65–71. doi:10.1093/annonc/mds227

Esteller, M., Silva, J.M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I.C., Repasky, E.A., Gabrielson, E., Schutte, M., Baylin, S.B., Herman, J.G., 2000. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J. Natl. Cancer Inst. 92, 564–569. doi:10.1093/jnci/92.7.564

Evans, D.G.R., 2009. Neurofibromatosis type 2 (NF2): a clinical and molecular review. Orphanet J. Rare Dis. 4, 16. doi:10.1186/1750-1172-4-16

Evans, P., Avey, S., Kong, Y., Krauthammer, M., 2013. Adjusting for background mutation frequency biases improves the identification of cancer driver genes. IEEE Trans. Nanobioscience 12, 150–7. doi:10.1109/TNB.2013.2263391

Exner, R., Pulverer, W., Diem, M., Spaller, L., Woltering, L., Schreiber, M., Wolf, B., Sonntagbauer, M., Schröder, F., Stift, J., Wrba, F., Bergmann, M., Weinhäusel, A., Egger, G., 2015. Potential of DNA methylation in rectal cancer as diagnostic and prognostic biomarkers. Br. J. Cancer 1–11. doi:10.1038/bjc.2015.303

Fackler, M.J., Umbricht, C.B., Williams, D., Argani, P., Cruz, L.-A., Merino, V.F., Teo, W.W., Zhang, Z., Huang, P., Visvananthan, K., Marks, J., Ethier, S., Gray, J.W., Wolff, A.C., Cope, L.M., Sukumar, S., 2011. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. Cancer Res. 71, 6195–207. doi:10.1158/0008-5472.CAN-11-1630

Faisal, K., Hospital, S., Arabia, S., Linda, L., Zealand, N., 1997. Editorial the Heidelberg Classification of Renal Cell. Am. J. Surg. Pathol. 133, 131–133.

Fausti, F., Di Agostino, S., Sacconi, A., Strano, S., Blandino, G., 2012. Hippo and rassf1a Pathways: A Growing Affair. Mol. Biol. Int. 2012, 307628. doi:10.1155/2012/307628

Feinberg, A.P., 2007. Phenotypic plasticity and the epigenetics of human disease. Nature 447, 433–40. doi:10.1038/nature05919

Felsenfeld, G., Groudine, M., 2003. Controlling the double helix. Nature 421, 448–53. doi:10.1038/nature01411

Feng, C., Sun, Y., Ding, G., Wu, Z., Jiang, H., Wang, L., Ding, Q., Wen, H., 2015. PI3Kβ Inhibitor TGX221 Selectively Inhibits Renal Cell Carcinoma Cells with Both VHL and SETD2 mutations and Links Multiple Pathways. Sci. Rep. 5, 9465. doi:10.1038/srep09465

Ferlay, J., Soerjomataram I, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman D, D., Bray, F., 2014. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer 136, E359–86. doi:10.1002/ijc.29210

Ferraro, A., 2016. Altered primary chromatin structures and their implications in cancer development. Cell. Oncol. doi:10.1007/s13402-016-0276-6

Ferrell, J.E., 2000. What do scaffold proteins really do? Sci. STKE 2000, pe1. doi:10.1126/stke.2000.52.pe1

Frattini, M., Ferrario, C., Bressan, P., Balestra, D., De Cecco, L., Mondellini, P., Bongarzone, I., Collini, P., Gariboldi, M., Pilotti, S., Pierotti, M.A., Greco, A., 2004. Alternative mutations of BRAF, RET and NTRK1 are associated with similar but distinct gene expression patterns in papillary thyroid cancer. Oncogene 23, 7436–7440. doi:10.1038/sj.onc.1207980

Freier, K., Knoepfle, K., Flechtenmacher, C., Pungs, S., Devens, F., Toedt, G., Hofele, C., Joos, S., Lichter, P., Radlwimmer, B., 2010. Recurrent copy number gain of transcription factor SOX2 and corresponding high protein expression in oral squamous cell carcinoma. Genes. Chromosomes Cancer 49, 9–16. doi:10.1002/gcc

Fridman, E., Dotan, Z., Barshack, I., David, M. Ben, Dov, A., Tabak, S., Zion, O., Benjamin, S., Benjamin, H., Kuker, H., Avivi, C., Rosenblatt, K., Polak-Charcon, S., Ramon, J., Rosenfeld, N., Spector, Y., 2010. Accurate molecular classification of renal tumors using microRNA expression. J. Mol. Diagn. 12, 687–96. doi:10.2353/jmoldx.2010.090187

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. U. S. A. 89, 1827–1831. doi:10.1073/pnas.89.5.1827

Fu, Z., Tindall, D.J., 2008. FOXOs, cancer and regulation of apoptosis. Oncogene 27, 2312–9. doi:10.1038/onc.2008.24

Gal-Yam, E.N., Saito, Y., Egger, G., Jones, P.A., 2008. Cancer epigenetics: modifications, screening, and therapy. Annu Rev Med 59, 267–280. doi:10.1146/annurev.med.59.061606.095816

Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., Ehrlich, M., 1983. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. 11, 6883–94.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. 6, pl1. doi:10.1126/scisignal.2004088

Garajová, I., Giovannetti, E., Biasco, G., Peters, G.J., 2015. c-Met as a Target for Personalized Therapy. Transl. Oncogenomics 7, 13–31. doi:10.4137/TOG.S30534

Ge, Y.-Z., Wu, R., Xin, H., Zhu, M., Lu, T.-Z., Liu, H., Xu, Z., Yu, P., Zhao, Y.-C., Li, M.-H., Hu, Z.-K., Zhao, Y., Zhong, B., Xu, X., Zhou, L.-H., Xu, L.-W., Wu, J.-P., Li, W.-C., Zhu, J.-G., Jia, R.-P., 2015. A tumor-specific microRNA signature predicts survival in clear cell renal cell carcinoma. J. Cancer Res. Clin. Oncol. 141, 1291–9. doi:10.1007/s00432-015-1927-0

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R., Martinez, P., Phillimore, B., Begum, S., Rabinowitz, A., Spencer-Dene, B., Gulati, S., Bates, P.A., Stamp, G., Pickering, L., Gore, M., Nicol, D.L., Hazell, S., Futreal, P.A., Stewart, A., Swanton, C., 2014. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat. Genet. 46, 225–33. doi:10.1038/ng.2891

Gibney, E.R., Nolan, C.M., 2010. Epigenetics and gene expression. Heredity (Edinb). 105, 4–13. doi:10.1038/hdy.2010.54

Gillespie, J.W., Ahram, M., Best, C.J., Swalwell, J.I., Krizman, D.B., Petricoin, E.F., Liotta, L.A., Emmert-Buck, M.R., 2001. The role of tissue microdissection in cancer research. Cancer J. 7, 32–9.

Giono, L.E., Manfredi, J.J., 2006. The p53 tumor suppressor participates in multiple cell cycle checkpoints. J. Cell. Physiol. 209, 13–20. doi:10.1002/jcp.20689

Gnarra, J.R., Tory, K., Weng, Y., Schmidt, L., Wei, M.H., Li, H., Latif, F., Liu, S., Chen, F., Duh, F.M., 1994. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nat. Genet. 7, 85–90. doi:10.1038/ng0594-85

Godin-Heymann, N., Wang, Y., Slee, E., Lu, X., 2013. Phosphorylation of ASPP2 by RAS/MAPK Pathway Is Critical for Its Full Pro-Apoptotic Function. PLoS One 8, e82022. doi:10.1371/journal.pone.0082022

Gossage, L., Eisen, T., Maher, E.R., 2015. VHL, the story of a tumour suppressor gene. Nat.

Publ. Gr. 15, 55–64. doi:10.1038/nrc3844

Grawenda, a M., O'Neill, E., 2015. Clinical utility of RASSF1A methylation in human malignancies. Br. J. Cancer 372–381. doi:10.1038/bjc.2015.221

Greaves, M., Maley, C.C., 2012. Clonal evolution in cancer. Nature 481, 306–313. doi:10.1038/nature10762

Greer, E.L., Brunet, A., 2005. FOXO transcription factors at the interface between longevity and tumor suppression. Oncogene 24, 7410–25. doi:10.1038/sj.onc.1209086

Gripp, K.W., Lin, A.E., 2012. Costello syndrome: a Ras/mitogen activated protein kinase pathway syndrome (rasopathy) resulting from HRAS germline mutations. Genet. Med. 14, 285–292. doi:10.1038/gim.0b013e31822dd91f

Guo, C., Tommasi, S., Liu, L., Yee, J.-K., Dammann, R., Pfeifer, G.P., 2007. RASSF1A Is Part of a Complex Similar to the Drosophila Hippo/Salvador/Lats Tumor-Suppressor Network. Curr. Biol. 17, 700–705. doi:10.1016/j.cub.2007.02.055

Guo, G., Sun, X., Chen, C., Wu, S., Huang, P., Li, Z., Dean, M., Huang, Y., Jia, W., Zhou, Q., Tang, A., Yang, Z., Li, X., Song, P., Zhao, X., Ye, R., Zhang, S., Lin, Z., Qi, M., Wan, S., Xie, L., Fan, F., Nickerson, M.L., Zou, X., Hu, X., Xing, L., Lv, Z., Mei, H., Gao, S., Liang, C., Gao, Z., Lu, J., Yu, Y., Liu, C., Li, L., Fang, X., Jiang, Z., Yang, J., Li, C., Zhao, X., Chen, J., Zhang, F., Lai, Y., Lin, Z., Zhou, F., Chen, H., Chan, H.C., Tsang, S., Theodorescu, D., Li, Y., Zhang, X., Wang, J., Yang, H., Gui, Y., Wang, J., Cai, Z., 2013. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. Nat. Genet. 45, 1459–63. doi:10.1038/ng.2798

Guo, J., Yang, Y., Yang, Y., Linghu, E., Zhan, Q., Brock, M. V, Herman, J.G., Zhang, B., Guo, M., 2015. RASSF10 suppresses colorectal cancer growth by activating P53 signaling and sensitizes colorectal cancer cell to docetaxel. Oncotarget 6, 4202–13.

Guo, Q., Wang, Y., Tripathi, P., Manda, K.R., Mukherjee, M., Chaklader, M., Austin, P.F., Surendran, K., Chen, F., 2015. Adam10 mediates the choice between principal cells and intercalated cells in the kidney. J. Am. Soc. Nephrol. 26, 149–59. doi:10.1681/ASN.2013070764

Guo, Y., Zhao, S., Sheng, Q., Ye, F., Li, J., Lehmann, B., Pietenpol, J., Samuels, D.C., Shyr, Y., 2014. Multi-perspective quality control of Illumina exome sequencing data using QC3. Genomics 103, 323–328. doi:10.1016/j.ygeno.2014.03.006

Haas, N.B., Lin, X., Manola, J., Pins, M., Liu, G., McDermott, D., Nanus, D., Heath, E., Wilding, G., Dutcher, J., 2012. A phase II trial of doxorubicin and gemcitabine in renal cell carcinoma with sarcomatoid features: ECOG 8802. Med. Oncol. 29, 761–7. doi:10.1007/s12032-011-9829-8

Habbig, S., Bartram, M.P., Müller, R.U., Schwarz, R., Andriopoulos, N., Chen, S., Sägmüller, J.G., Hoehne, M., Burst, V., Liebau, M.C., Reinhardt, H.C., Benzing, T., Schermer, B., 2011. NPHP4, a cilia-associated protein, negatively regulates the Hippo pathway. J. Cell Biol. 193, 633–642. doi:10.1083/jcb.201009069

Hagenkord, J.M., Gatalica, Z., Jonasch, E., Monzon, F.A., 2011. Clinical genomics of renal epithelial tumors. Cancer Genet. 204, 285–97. doi:10.1016/j.cancergen.2011.06.001

Hagrass, H.A., Pasha, H.F., Shaheen, M.A., Abdel Bary, E.H., Kassem, R., 2013. Methylation status and protein expression of RASSF1A in breast cancer patients. Mol. Biol. Rep. 41, 57–65. doi:10.1007/s11033-013-2837-3

Hakimi, a A., Pham, C.G., Hsieh, J.J., 2013. A clear picture of renal cell carcinoma. Nat. Genet. 45, 849–50. doi:10.1038/ng.2708

Hakimi, A.A., Ostrovnaya, I., Reva, B., Schultz, N., Chen, Y.-B., Gonen, M., Liu, H., Takeda, S., Voss, M.H., Tickoo, S.K., Reuter, V.E., Russo, P., Cheng, E.H., Sander, C., Motzer, R.J., Hsieh, J.J., 2013. Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. Clin.

Cancer Res. 19, 3259–67. doi:10.1158/1078-0432.CCR-12-3886

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144, 646–74. doi:10.1016/j.cell.2011.02.013

Harb-De la Rosa, A., Acker, M., Swain, S., Manoharan, M., 2015. The role of epigenetics in kidney malignancies. Cent. Eur. J. Urol. 68. doi:10.5173/ceju.2015.453

Harper, K.N., Peters, B.A., Gamble, M. V, 2013. Batch effects and pathway analysis: two potential perils in cancer studies involving DNA methylation array analysis. Cancer Epidemiol. Biomarkers Prev. 22, 1052–60. doi:10.1158/1055-9965.EPI-13-0114

Harris, S.L., Levine, A.J., 2005. The p53 pathway: positive and negative feedback loops. Oncogene 24, 2899–2908. doi:10.1038/sj.onc.1208615

Harvey, K.F., Zhang, X., Thomas, D.M., 2013. The Hippo pathway and human cancer. Nat Rev Cancer 13, 246–257. doi:nrc3458 [pii]\n10.1038/nrc3458

Hayatsu, H., 2008. Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis--a personal account. Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci. 84, 321–30.

Henrique, R., Jer, C., 2015. New insights on chromatin modifiers and histone post-translational modifications in renal cell tumours 16–24. doi:10.1111/eci.12360

Henrique, R., Luís, A.S., Jerónimo, C., 2012. The epigenetics of renal cell tumors: From biology to biomarkers. Front. Genet. 3, 1–13. doi:10.3389/fgene.2012.00094

Herman, J.G., Graff, J.R., Myöhänen, S., Nelkin, B.D., Baylin, S.B., 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc. Natl. Acad. Sci. U. S. A. 93, 9821–6.

Herman, J.G., Latif, F., Weng, Y., Lerman, M.I., Zbar, B., Liu, S., Samid, D., Duan, D.S., Gnarra, J.R., Linehan, W.M., 1994. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc. Natl. Acad. Sci. U. S. A. 91, 9700–9704. doi:10.1073/pnas.91.21.9700

Hesson, L., Bièche, I., Krex, D., Criniere, E., Hoang-Xuan, K., Maher, E.R., Latif, F., 2004. Frequent epigenetic inactivation of RASSF1A and BLU genes located within the critical 3p21.3 region in gliomas. Oncogene 23, 2408–2419. doi:10.1038/sj.onc.1207407

Hesson, L.B., Cooper, W.N., Latif, F., 2007. The role of RASSF1A methylation in cancer. Dis. Markers 23, 73–87.

Hesson, L.B., Dunwell, T.L., Cooper, W.N., Catchpoole, D., Brini, A.T., Chiaramonte, R., Griffiths, M., Chalmers, A.D., Maher, E.R., Latif, F., 2009. The novel RASSF6 and RASSF10 candidate tumour suppressor genes are frequently epigenetically inactivated in childhood leukaemias. Mol. Cancer 8, 42. doi:10.1186/1476-4598-8-42

Heyn, H., Esteller, M., 2012. DNA methylation profiling in the clinic: applications and challenges. Nat. Rev. Genet. 13, 679–692. doi:10.1038/nrg3270

Heyn, H., Moran, S., Hernando-Herraez, I., Sayols, S., Gomez, A., Sandoval, J., Monk, D., Hata, K., Marques-Bonet, T., Wang, L., Esteller, M., 2013. DNA methylation contributes to natural human variation. Genome Res. 23, 1363–72. doi:10.1101/gr.154187.112

Heyn, H., Vida, E., Sayols, S., Sanchez-Mut, J. V., Moran, S., Medina, I., Sandoval, J., Simó-Riudalbas, L., Szczesna, K., Huertas, D., Gatto, S., Matarazzo, M.R., Dopazo, J., Esteller, M., 2012. Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. Epigenetics 7, 542–550. doi:10.4161/epi.20523

Hill, V.K., Dunwell, T.L., Catchpoole, D., Krex, D., Brini, A.T., Griffiths, M., Craddock, C., Maher, E.R., Latif, F., 2011. Frequent epigenetic inactivation of KIBRA, an upstream member of the Salvador/Warts/Hippo (SWH) tumor suppressor network, is associated with specific genetic event in B-cell acute lymphocytic leukemia. Epigenetics 6, 326–32.

Hill, V.K., Ricketts, C., Bieche, I., Vacher, S., Gentle, D., Lewis, C., Maher, E.R., Latif, F., 2011.

Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity. Cancer Res. 71, 2988–2999. doi:10.1158/0008-5472.CAN-10-4026

Hill, V.K., Underhill-Day, N., Krex, D., Robel, K., Sangan, C.B., Summersgill, H.R., Morris, M., Gentle, D., Chalmers, A.D., Maher, E.R., Latif, F., 2011. Epigenetic inactivation of the RASSF10 candidate tumor suppressor gene is a frequent and an early event in gliomagenesis. Oncogene 30, 978–89. doi:10.1038/onc.2010.471

Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C., Kitano, T.K., Hodel, M.R., Petersen, J.F., Wyatt, P.W., Steenblock, E.R., Shah, P.H., Bousse, L.J., Troup, C.B., Mellen, J.C., Wittmann, D.K., Erndt, N.G., Cauley, T.H., Koehler, R.T., So, A.P., Dube, S., Rose, K.A., Montesclaros, L., Wang, S., Stumbo, D.P., Hodges, S.P., Romine, S., Milanovich, F.P., White, H.E., Regan, J.F., Karlin-Neumann, G.A., Hindson, C.M., Saxonov, S., Colston, B.W., 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal. Chem. 83, 8604–10. doi:10.1021/ac202028g

Hirakawa, M., 2002. JSNP: a database of common gene variations in the Japanese population. Nucleic Acids Res. 30, 158–162. doi:10.1093/nar/30.1.158

Ho, T.H., Kapur, P., Joseph, R.W., Serie, D.J., Eckel-Passow, J.E., Parasramka, M., Cheville, J.C., Wu, K.J., Frenkel, E., Rakheja, D., Stefanius, K., Brugarolas, J., Parker, A.S., 2015. Loss of PBRM1 and BAP1 expression is less common in non-clear cell renal cell carcinoma than in clear cell renal cell carcinoma. Urol. Oncol. 33, 23.e9–14. doi:10.1016/j.urolonc.2014.10.014

Hoffman, A.M., Cairns, P., 2011. Epigenetics of kidney cancer and bladder cancer. Epigenomics 3, 19–34. doi:10.2217/epi.10.64

Hong, X., Nguyen, H.T., Chen, Q., Zhang, R., Hagman, Z., Mathijs, P., Cohen, S.M., 2014. Opposing activities of the Ras and Hippo pathways converge on regulation of YAP protein turnover 33, 2447–2457.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., Schadendorf, D., Kumar, R., 2013. TERT promoter mutations in familial and sporadic melanoma. Science 339, 959–61. doi:10.1126/science.1230062

Hosen, I., Rachakonda, P.S., Heidenreich, B., Sitaram, R.T., Ljungberg, B., Roos, G., Hemminki, K., Kumar, R., 2015. TERT promoter mutations in clear cell renal cell carcinoma. Int. J. Cancer 136, 2448–52. doi:10.1002/ijc.29279

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57. doi:10.1038/nprot.2008.211

Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G. V, Chin, L., Garraway, L.A., 2013. Highly recurrent TERT promoter mutations in human melanoma. Science 339, 957–9. doi:10.1126/science.1229259

Humphrey, P.A., 2012. Sarcomatoid renal cell carcinoma. J. Urol. 188, 601–2. doi:10.1016/j.juro.2012.05.021

Hunt, J.D., van der Hel, O.L., McMillan, G.P., Boffetta, P., Brennan, P., 2005. Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. Int. J. Cancer 114, 101–8. doi:10.1002/ijc.20618

Hwang, E., Cheong, H.-K., Ul Mushtaq, A., Kim, H.-Y., Yeo, K.J., Kim, E., Lee, W.C., Hwang, K.Y., Cheong, C., Jeon, Y.H., 2014. Structural basis of the heterodimerization of the MST and RASSF SARAH domains in the Hippo signalling pathway. Acta Crystallogr. D. Biol. Crystallogr. 70, 1944–53. doi:10.1107/S139900471400947X

Ibragimova, I., Maradeo, M.E., Dulaimi, E., Cairns, P., 2013a. Aberrant promoter hypermethylation of PBRM1, BAP1, SETD2, KDM6A and other chromatin-modifying genes is

absent or rare in clear cell RCC. Epigenetics 8, 486–93. doi:10.4161/epi.24552

Ibragimova, I., Slifker, M.J., Maradeo, M.E., Banumathy, G., Dulaimi, E., Uzzo, R.G., Cairns, P., 2013b. Genome-Wide Promoter Methylome of Small Renal Masses. PLoS One 8, e77309. doi:10.1371/journal.pone.0077309

Illumina, 2012. HumanMethylation450 BeadChip Achieves Breadth of Coverage Using Two Infinium ® Chemistries. Tech. Note Epigenetic Anal. 2–5.

Illumina, 2006. GoldenGate® Assay for Methylation and BeadArrayTM Technology. Tech. Note Epigenetic Anal.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J.B., Sabunciyan, S., Feinberg, A.P., 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. 41, 178–86. doi:10.1038/ng.298

Ishmael, F.T., Stellato, C., Mullis, K., Faloona, F., Mullis, K., Olson, M., Hood, L., Cantor, C., Botstein, D., Panet, A., Khorana, H., Wittwer, C., Herrmann, M., Moss, A., Rasmussen, R., Palmer, S., Wiegand, A., Maldarelli, F., al., et, Saiki, R., Scharf, S., Faloona, F., al., et, Hentemann, M., Reiss, J., Wagner, M., Cooper, D., Hensel, C., Xiang, R., Sakaguchi, A., Naylor, S., McGee, T., Cowley, G., Yandell, D., Dryja, T., Tamaki, K., Jeffreys, A., Tang, Y., Procop, G., Persing, D., Mackay, I., Wong, M., Medrano, J., Valasek, M., Repa, J., Gibson, U., Heid, C., Williams, P., Clegg, R., Murchie, A., Zechel, A., Carlberg, C., Diekmann, S., Lilley, D., Tichopad, A., Dilger, M., Schwarz, G., Pfaffl, M., Kuhne, B., Oschmann, P., Baliga, C., Sutton, R., 2008. Principles and applications of polymerase chain reaction: basic science for the practicing physician. Ann. Allergy, Asthma Immunol. 101, 437–443. doi:10.1016/S1081-1206(10)60323-7

Issa, J.-P., 2004. CpG island methylator phenotype in cancer. Nat. Rev. Cancer 4, 988–993. doi:10.1038/nrc1507

Iwasa, H., Kudo, T., Maimaiti, S., Ikeda, M., Maruyama, J., Nakagawa, K., Hata, Y., 2013. The RASSF6 tumor suppressor protein regulates apoptosis and the cell cycle via MDM2 protein and p53 protein. J. Biol. Chem. 288, 30320–9. doi:10.1074/jbc.M113.507384

Jafri, M., Wake, N.C., Ascher, D.B., Pires, D.E. V, Gentle, D., Morris, M.R., Rattenberry, E., Simpson, M.A., Trembath, R.C., Weber, A., Woodward, E.R., Donaldson, A., Blundell, T.L., Latif, F., Maher, E.R., 2015. Germline Mutations in the CDKN2B tumor suppressor gene predispose to renal cell carcinoma. Cancer Discov. CD–14–1096. doi:10.1158/2159-8290.CD-14-1096

Jamuar, S.S., Tan, E.-C., Foundation, M., Baird, P., Anderson, T., et al., 2015. Clinical application of next-generation sequencing for Mendelian diseases. Hum. Genomics 9, 10. doi:10.1186/s40246-015-0031-5

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., von Mering, C., 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37, D412–6. doi:10.1093/nar/gkn760

Jiang, F., Moch, H., Richter, J., Egenter, C., Gasser, T., Bubendorf, L., Gschwind, R., Sauter, G., Mihatsch, M.J., 1998. Comparative genomic hybridization reveals frequent chromosome 13q and 4q losses in renal carcinomas with sarcomatoid transformation. J. Pathol. 185, 382–388. doi:10.1002/(SICI)1096-9896(199808)185:4<382::AID-PATH124>3.0.CO;2-4

Jin, S.-G., Kadam, S., Pfeifer, G.P., 2010. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 38, e125. doi:10.1093/nar/gkq223

Jing, L., Su, L., Ring, B.Z., 2014. Ethnic background and genetic variation in the evaluation of cancer risk: a systematic review. PLoS One 9, e97522. doi:10.1371/journal.pone.0097522

Johansson, M.E., Axelson, H., Linehan, W.M., Ricketts, C.J., 2013. The metabolic basis of kidney

cancer. Semin. Cancer Biol. 23, 46–55.

John, E.M., Miron, A., Gong, G., Phipps, A.I., Felberg, A., Li, F.P., West, D.W., Whittemore, A.S., 2007. Prevalence of pathogenic BRCA1 mutation carriers in 5 US racial/ethnic groups. JAMA 298, 2869–76. doi:10.1016/S0084-3954(08)79042-0

Johnson, R., Halder, G., 2014. The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment. Nat. Rev. Drug Discov. 13, 63–79. doi:10.1038/nrd4161

Jonasch, E., Gao, J., Rathmell, W.K., 2014. Renal cell carcinoma. BMJ 349, g4797. doi:10.1136/bmj.g4797

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–92. doi:10.1038/nrg3230

Jones, T.D., Eble, J.N., Wang, M., MacLennan, G.T., Jain, S., Cheng, L., 2005. Clonal divergence and genetic heterogeneity in clear cell renal cell carcinomas with sarcomatoid transformation. Cancer 104, 1195–1203. doi:10.1002/cncr.21288

Joshi, S., Tolkunov, D., Aviv, H., Hakimi, A.A., Yao, M., Hsieh, J.J., Ganesan, S., Chan, C.S., White, E., 2015. The Genomic Landscape of Renal Oncocytoma Identifies a Metabolic Barrier to Tumorigenesis. Cell Rep. 13, 1895–1908. doi:10.1016/j.celrep.2015.10.059

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., Crabtree, G.R., 2013. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat. Genet. advance on. doi:10.1038/ng.2628

Kalari, S., Pfeifer, G.P., 2010. Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. Adv. Genet. 70, 277–308. doi:10.1016/B978-0-12-380866-0.60010-1

Kanamaru, H., Li, B., Miwa, Y., Akino, H., Okada, K., 1999. Immunohistochemical expression of p53 and bcl-2 proteins is not associated with sarcomatoid change in renal cell carcinoma. Urol. Res. 27, 169–173.

Kapur, P., Peña-Llopis, S., Christie, A., Zhrebker, L., Pavía-Jiménez, A., Rathmell, W.K., Xie, X.J., Brugarolas, J., 2013. Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: A retrospective analysis with independent validation. Lancet Oncol. 14, 159–167. doi:10.1016/S1470-2045(12)70584-3

Karakas, B., Bachman, K.E., Park, B.H., 2006. Mutation of the PIK3CA oncogene in human cancers. Br. J. Cancer 94, 455–9. doi:10.1038/sj.bjc.6602970

Karami, S., Andreotti, G., Liao, L.M., Pfeiffer, R.M., Weinstein, S.J., Purdue, M.P., Hofmann, J.N., Albanes, D., Mannisto, S., Moore, L.E., 2015. LINE1 methylation levels in pre-diagnostic leukocyte DNA and future renal cell carcinoma risk. Epigenetics 10, 282–292. doi:10.1080/15592294.2015.1006505

Karnoub, A.E., Weinberg, R.A., 2008. Ras oncogenes: split personalities. Nat. Rev. Mol. Cell Biol. 9, 517–31. doi:10.1038/nrm2438

Katoh, M., Katoh, M, 2004. Human FOX gene family (Review). Int. J. Oncol. 25, 1495–1500.

Katoh, M.M.M.M., Igarashi, M., Fukuda, H., Nakagama, H., Katoh, M.M.M.M., 2013. Cancer genetics and genomics of human FOX family genes. Cancer Lett. 328, 198–206. doi:10.1016/j.canlet.2012.09.017

Kendall, B.J., Wilson, L.F., Olsen, C.M., Webb, P.M., Neale, R.E., Bain, C.J., Whiteman, D.C., 2015. Cancers in Australia in 2010 attributable to overweight and obesity. Aust. N. Z. J. Public Health 39, 452–7. doi:10.1111/1753-6405.12458

Kim, S.Y., Speed, T.P., 2013. Comparing somatic mutation-callers: beyond Venn diagrams. BMC Bioinformatics 14, 189. doi:10.1186/1471-2105-14-189

Kim, T., Zargar-Shoshtari, K., Dhillon, J., Lin, H.-Y., Yue, B., Fishman, M., Sverrisson, E.F., Spiess, P.E., Gupta, S., Poch, M.A., Sexton, W.J., 2015. Using Percentage of Sarcomatoid Differentiation

as a Prognostic Factor in Renal Cell Carcinoma. Clin. Genitourin. Cancer 13, 225–230. doi:10.1016/j.clgc.2014.12.001

King, R.J.B., Robins, M.W., 2006. Cancer Biology. Pearson/Prentice Hall.

Klatte, T., Pantuck, A.J., Said, J.W., Seligson, D.B., Rao, N.P., LaRochelle, J.C., Shuch, B., Zisman, A., Kabbinavar, F.F., Belldegrun, A.S., 2009. Cytogenetic and molecular tumor profiling for type 1 and type 2 papillary renal cell carcinoma. Clin. Cancer Res. 15, 1162–9. doi:10.1158/1078-0432.CCR-08-1229

Kluzek, K., Bluyssen, H.A., Wesoly, J., 2015. The epigenetic landscape of clear-cell renal cell carcinoma. J. Kidney Cancer VHL 2, 90. doi:10.15586/jkcvhl.2015.33

Knudson, a G., 1971. Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. U. S. A. 68, 820–823. doi:10.1073/pnas.68.4.820

Knudson, A.G., 2001. Two genetic hits (more or less) to cancer. Nat. Rev. Cancer 1, 157–162. doi:10.1038/35101031

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–76. doi:10.1101/gr.129684.111

Koelsche, C., Renner, M., Hartmann, W., Brandt, R., Lehner, B., Waldburger, N., Alldinger, I., Schmitt, T., Egerer, G., Penzel, R., Wardelmann, E., Schirmacher, P., von Deimling, A., Mechtersheimer, G., 2014. TERT promoter hotspot mutations are recurrent in myxoid liposarcomas but rare in other soft tissue sarcoma entities. J. Exp. Clin. Cancer Res. 33, 33. doi:10.1186/1756-9966-33-33

Kolch, W., 2005. Coordinating ERK/MAPK signalling through scaffolds and inhibitors. Nat. Rev. Mol. Cell Biol. 6, 827–837. doi:10.1038/nrm1743

Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23, 1289–1291. doi:10.1093/bioinformatics/btm091

Kouzarides, T., 2007. Chromatin modifications and their function. Cell 128, 693–705. doi:10.1016/j.cell.2007.02.005

Kovacs, G., Akhtar, M., Beckwith, B.J., Bugert, P., Cooper, C.S., Delahunt, B., Eble, J.N., Fleming, S., Ljungberg, B., Medeiros, L.J., Moch, H., Reuter, V.E., Ritz, E., Roos, G., Schmidt, D., Srigley, J.R., Störkel, S., van den Berg, E., Zbar, B., 1997. The Heidelberg classification of renal cell tumours. J. Pathol. 183, 131–3. doi:10.1002/(SICI)1096-9896(199710)183:2<131::AID-PATH931>3.0.CO;2-G

Ku, C., Cooper, D.N., Polychronakos, C., Naidoo, N., Wu, M., Soong, R., 2012. Exome Sequencing : Dual Role as a Discovery and Diagnostic Tool 5–14. doi:10.1002/ana.22647

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–81. doi:10.1038/nprot.2009.86

Kvasha, S., Gordiyuk, V., Kondratov, A., Ugryn, D., Zgonnyk, Y.M., Rynditch, A. V., Vozianov, A.F., 2008. Hypermethylation of the 5'CpG island of the FHIT gene in clear cell renal carcinomas. Cancer Lett. 265, 250–257. doi:10.1016/j.canlet.2008.02.036

Lando, M., Fjeldbo, C.S., Wilting, S.M., Snoek, B., Aarnes, E.-K., Forsberg, M.F., Kristensen, G.B., Steenbergen, R.D.M., Lyng, H., 2015. Interplay between promoter methylation and chromosomal loss in gene silencing at 3p11-p14 in cervical cancer. Epigenetics. doi:10.1080/15592294.2015.1085140

Lang, M., Vocke, C.D., Merino, M.J., Schmidt, L.S., Linehan, W.M., 2015. Mitochondrial DNA mutations distinguish bilateral multifocal renal oncocytomas from familial Birt-Hogg-Dubé tumors. Mod. Pathol. 28, 1458–69. doi:10.1038/modpathol.2015.101

Langton, P.F., Colombani, J., Chan, E.H.Y., Wepf, A., Gstaiger, M., Tapon, N., 2009. The dASPP-dRASSF8 Complex Regulates Cell-Cell Adhesion during Drosophila Retinal Morphogenesis. Curr. Biol. 19, 1969–1978. doi:10.1016/j.cub.2009.10.027

Lasseigne, B.N., Burwell, T.C., Patil, M. a, Absher, D.M., Brooks, J.D., Myers, R.M., 2014. DNA methylation profiling reveals novel diagnostic biomarkers in renal cell carcinoma. BMC Med. 12, 235. doi:10.1186/PREACCEPT-1273016012136123

Latif, F., Tory, K., Gnarra, J., Yao, M., Duh, F.M., Orcutt, M.L., Stackhouse, T., Kuzmin, I., Modi, W., Geil, L., 1993. Identification of the von Hippel-Lindau disease tumor suppressor gene. Science 260, 1317–1320. doi:10.1002/anie.201102909

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G. V, Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S. a, Kiezun, A., Hammerman, P.S., McKenna, A., Drier, Y., Zou, L., Ramos, A.H., Pugh, T.J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortés, M.L., Auclair, D., Saksena, G., Voet, D., Noble, M., DiCara, D., Lin, P., Lichtenstein, L., Heiman, D.I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A.M., Lohr, J., Landau, D.-A., Wu, C.J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. a, Mora, J., Lee, R.S., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S.B., Roberts, C.W.M., Biegel, J. a, Stegmaier, K., Bass, A.J., Garraway, L. a, Meyerson, M., Golub, T.R., Gordenin, D.A., Sunyaev, S., Lander, E.S., Getz, G., 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214–8. doi:10.1038/nature12213

Ledford, H., 2013. Lists of cancer mutations awash with false positives. Nature. doi:10.1038/nature.2013.13206

Lee, E.Y.H.P., Muller, W.J., 2010. Oncogenes and tumor suppressor genes. Cold Spring Harb. Perspect. Biol. 2, a003236. doi:10.1101/cshperspect.a003236

Lee, J., Seol, M.-Y., Jeong, S., Kwon, H.J.J., Lee, C.R., Ku, C.R., Kang, S.-W.S.-W., Jeong, J.J., Shin, D.Y., Nam, K.-H.K.-H., Lee, E.J., Chung, W.Y., Jo, Y.S., 2015. KSR1 is coordinately regulated with Notch signaling and oxidative phosphorylation in thyroid cancer. J. Mol. Endocrinol. 54, 115–24. doi:10.1530/JME-14-0270

Lee, W.H., Bookstein, R., Hong, F., Young, L.J., Shew, J.Y., Lee, E.Y., 1987. Human retinoblastoma susceptibility gene: cloning, identification, and sequence. Science 235, 1394–9.

Lelieveld, S.H., Spielmann, M., Mundlos, S., Veltman, J.A., Gilissen, C., 2015. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. doi:10.1002/humu.22813

Leroy, B., Anderson, M., Soussi, T., 2014. TP53 mutations in human cancer: database reassessment and prospects for the next decade. Hum. Mutat. 35, 672–88. doi:10.1002/humu.22552

Li, B., Lu, L., Zhong, M., Tan, X.X., Liu, C.Y., Guo, Y., Yi, X., 2013. Terbinafine inhibits KSR1 and suppresses Raf-MEK-ERK signaling in oral squamous cell carcinoma cells. Neoplasma 60, 406–12. doi:10.4149/neo_2013_052

Li, M.-Y., Xu, Y.-Y., Kang, H.-Y., Wang, X.-R., Gao, L., Cen, J., Wang, W., Wang, N., Li, Y.-H., Wang, L.-L., Yu, L., 2015. Quantitative Detection of ID4 Gene Aberrant Methylation in the Differentiation of Myelodysplastic Syndrome from Aplastic Anemia. Chin. Med. J. (Engl). 128, 2019–2025. doi:10.4103/0366-6999.161351

Li, Y., Tollefsbol, T.O., 2011. DNA methylation detection: bisulfite genomic sequencing analysis. Methods Mol. Biol. 791, 11–21. doi:10.1007/978-1-61779-316-5_2

Li, Z., Chang, X., Dai, D., Deng, P., Sun, Q., 2014. RASSF10 is an epigenetically silenced tumor suppressor in gastric cancer. Oncol. Rep. 31, 1661–8. doi:10.3892/or.2014.3039

Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., Lee, C.-

W., Hu, D., Lian, B.Q., Kleffel, S., Yang, Y., Neiswender, J., Khorasani, A.J., Fang, R., Lezcano, C., Duncan, L.M., Scolyer, R.A., Thompson, J.F., Kakavand, H., Houvras, Y., Zon, L.I., Mihm, M.C., Kaiser, U.B., Schatton, T., Woda, B.A., Murphy, G.F., Shi, Y.G., 2012. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell 150, 1135–46. doi:10.1016/j.cell.2012.07.033

Lienkamp, S., Ganner, A., Walz, G., 2012. Inversin, Wnt signaling and primary cilia. Differentiation 83, S49–S55. doi:10.1016/j.diff.2011.11.012

Lind, G.E., Raiborg, C., Danielsen, S. a, Rognum, T.O., Thiis-Evensen, E., Hoff, G., Nesbakken, a, Stenmark, H., Lothe, R. a, 2011. SPG20, a novel biomarker for early detection of colorectal cancer, encodes a regulator of cytokinesis. Oncogene 30, 3967–3978. doi:10.1038/onc.2011.109

Lindqvist, B.M., Wingren, S., Motlagh, P.B., Nilsson, T.K., 2014. Whole genome DNA methylation signature of HER2-positive breast cancer. Epigenetics 9, 1149–62. doi:10.4161/epi.29632

Linehan, W.M., Pinto, P.A., Bratslavsky, G., Pfaffenroth, E., Merino, M., Vocke, C.D., Toro, J.R., Bottaro, D., Neckers, L., Schmidt, L.S., Srinivasan, R., 2009. Hereditary kidney cancer: unique opportunity for disease-based therapy. Cancer 115, 2252–61. doi:10.1002/cncr.24230

Linehan, W.M., Ricketts, C.J., 2014. Decade in review—kidney cancer: Discoveries, therapies and opportunities. Nat. Rev. Urol. 11, 614–616. doi:10.1038/nrurol.2014.262

Linehan, W.M., Srinivasan, R., Schmidt, L.S., 2010. The genetic basis of kidney cancer: a metabolic disease. Nat. Rev. Urol. 7, 277–85. doi:10.1038/nrurol.2010.47

Liu, Q., Su, P.-F., Zhao, S., Shyr, Y., 2014. Transcriptome-wide signatures of tumor stage in kidney renal clear cell carcinoma: connecting copy number variation, methylation and transcription factor activity. Genome Med. 6, 117. doi:10.1186/s13073-014-0117-z

Lopez-Beltran, A., Carrasco, J.C., Cheng, L., Scarpelli, M., Kirkali, Z., Montironi, R., 2009. 2009 Update on the Classification of Renal Epithelial Tumors in Adults. Int. J. Urol. 16, 432–443. doi:10.1111/j.1442-2042.2009.02302.x

Lopez-Beltran, A., Montironi, R., Egevad, L., Caballero-Vargas, M.T., Scarpelli, M., Kirkali, Z., Cheng, L., 2010. Genetic profiles in renal tumors: Review Article. Int. J. Urol. 17, 6–19. doi:10.1111/j.1442-2042.2009.02395.x

Lopez-Beltran, A., Scarpelli, M., Montironi, R., Kirkali, Z., 2006. 2004 WHO classification of the renal tumors of the adults. Eur. Urol. 49, 798–805. doi:10.1016/j.eururo.2005.11.035

Lopez-Serra, P., Esteller, M., 2012. DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. Oncogene 31, 1609–22. doi:10.1038/onc.2011.354

Lorenz, T.C., 2012. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. J. Vis. Exp. e3998. doi:10.3791/3998

Lorincz, M.C., Dickerson, D.R., Schmitt, M., Groudine, M., 2004. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat. Struct. Mol. Biol. 11, 1068–1075. doi:10.1038/nsmb840

Lovell, M., Lott, S.T., Wong, P., El-Naggar, A., Tucker, S., Killary, A.M., 1999. The genetic locus NRC-1 within chromosome 3p12 mediates tumor suppression in renal cell carcinoma independently of histological type, tumor microenvironment, and VHL mutation. Cancer Res. 59, 2182–9.

Lujambio, a, Portela, a, Liz, J., Melo, S. a, Rossi, S., Spizzo, R., Croce, C.M., Calin, G. a, Esteller, M., 2010. CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. Oncogene 29, 6390–6401. doi:10.1038/onc.2010.361

Maher, E.R., 2013. Genomics and epigenomics of renal cell carcinoma. Semin. Cancer Biol. doi:10.1016/j.semcancer.2012.06.003

Maher, E.R., 2011. Genetics of familial renal cancers. Nephron - Exp. Nephrol. 118, 21–26. doi:10.1159/000320892

Maksimovic, J., Gordon, L., Oshlack, A., 2012. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol. 13, R44. doi:10.1186/gb-2012-13-6-r44

Malone, K.E., Daling, J.R., Doody, D.R., Hsu, L., Bernstein, L., Coates, R.J., Marchbanks, P. a., Simon, M.S., McDonald, J. a., Norman, S. a., Strom, B.L., Burkman, R.T., Ursin, G., Deapen, D., Weiss, L.K., Folger, S., Madeoy, J.J., Friedrichsen, D.M., Suter, N.M., Humphrey, M.C., Spirtas, R., Ostrander, E. a., 2006. Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in White and Black American women ages 35 to 64 years. Cancer Res. 66, 8297–8308. doi:10.1158/0008-5472.CAN-06-0503

Malouf, G.G., Su, X., Yao, H., Gao, J., Xiong, L., He, Q., Compérat, E., Couturier, J., Molinié, V., Escudier, B., Camparo, P., Doss, D.J., Thompson, E.J., Khayat, D., Wood, C.G., Yu, W., Teh, B.T., Weinstein, J., Tannir, N.M., 2014. Next-generation sequencing of translocation renal cell carcinoma reveals novel RNA splicing partners and frequent mutations of chromatin-remodeling genes. Clin. Cancer Res. 20, 4129–40. doi:10.1158/1078-0432.CCR-13-3036

Malzkorn, B., Wolter, M., Riemenschneider, M.J., Reifenberger, G., 2011. Unraveling the glioma epigenome: from molecular mechanisms to novel biomarkers and therapeutic targets. Brain Pathol. 21, 619–32. doi:10.1111/j.1750-3639.2011.00536.x

Mamedov, T.G., Pienaar, E., Whitney, S.E., TerMaat, J.R., Carvill, G., Goliath, R., Subramanian, A., Viljoen, H.J., 2008. A fundamental study of the PCR amplification of GC-rich DNA templates. Comput. Biol. Chem. 32, 452–7. doi:10.1016/j.compbiolchem.2008.07.021

Marsaud, A., Dadone, B., Ambrosetti, D., Baudoin, C., Chamorey, E., Rouleau, E., Lefol, C., Roussel, J.-F., Fabas, T., Cristofari, G., Carpentier, X., Michiels, J.-F., Amiel, J., Pedeutour, F., 2015. Dismantling papillary renal cell carcinoma classification: The heterogeneity of genetic profiles suggests several independent diseases. Genes. Chromosomes Cancer 54, 369–82. doi:10.1002/gcc.22248

Matallanas, D., Romano, D., Yee, K., Meissl, K., Kucerova, L., Piazzolla, D., Baccarini, M., Vass, J.K., Kolch, W., O'neill, E., 2007. RASSF1A elicits apoptosis through an MST2 pathway directing proapoptotic transcription by the p73 tumor suppressor protein. Mol. Cell 27, 962–75. doi:10.1016/j.molcel.2007.08.008

McGranahan, N., Swanton, C., 2015. Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. Cancer Cell 27, 15–26. doi:10.1016/j.ccell.2014.12.001

McRonald, F.E., Morris, M.R., Gentle, D., Winchester, L., Baban, D., Ragoussis, J., Clarke, N.W., Brown, M.D., Kishida, T., Yao, M., Latif, F., Maher, E.R., 2009. CpG methylation profiling in VHL related and VHL unrelated renal cell carcinoma. Mol. Cancer 8, 31. doi:10.1186/1476-4598-8-31

Meienberg, J., Zerjavic, K., Keller, I., Okoniewski, M., Patrignani, A., Ludin, K., Xu, Z., Steinmann, B., Carrel, T., Röthlisberger, B., Schlapbach, R., Bruggmann, R., Matyas, G., 2015. New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res. 43, e76. doi:10.1093/nar/gkv216

Meserve, J.H., Duronio, R.J., 2015. Scalloped and Yorkie are required for cell cycle re-entry of quiescent cells after tissue damage. Development 2740–2751. doi:10.1242/dev.119339

Metzker, M.L., 2010. Sequencing technologies - the next generation. Nat. Rev. Genet. 11, 31–46. doi:10.1038/nrg2626

Meynert, A.M., Ansari, M., FitzPatrick, D.R., Taylor, M.S., 2014. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics 15, 247. doi:10.1186/1471-2105-15-247

Mian, B.M., Bhadkamkar, N., Slaton, J.W., Tomera, K.M., Farrow, G.M., Lieber, M.M., Ro, J.Y.,

Ayala, A.G., Sella, A., Al., E., Sella, A., Logothetis, C.J., Ro, J.Y., Al., E., Bertoni, F., Ferri, C., Benati, A., Al., E., Oda, H., Machinami, R., Culine, S., Bekradda, M., Terrier-Lacombe, M.J., Al., E., Cangiano, T., Liao, J., Naitoh, J., Al., E., Cancer, A.J.C. on, Fuhrman, S.A., Lasky, L.C., Limas, C., Flanigan, R.C., Blumenstein, B.A., Salmon, S., Al., E., Mickisch, G.H., Garin, A., Madej, M., Al., E., Krutchik, A.N., Sullivan, C., Sinkovics, J.G., Al., E., Motzer, R.J., Mazumdar, M., Bacik, J., Al., E., Kovacs, G., Akhtar, M., Beckwith, B.J., Al., E., Storkel, S., Eble, J.N., Adlakha, K., Al., E., Farrow, G.M., Harrison, E.G., Utz, D.C., 2002. Prognostic factors and survival of patients with sarcomatoid renal cell carcinoma. J. Urol. 167, 65–70. doi:10.1016/S0022-5347(05)65384-0

Michaelson, M.D., McKay, R.R., Werner, L., Atkins, M.B., Van Allen, E.M., Olivier, K.M., Song, J., Signoretti, S., McDermott, D.F., Choueiri, T.K., 2015. Phase 2 trial of sunitinib and gemcitabine in patients with sarcomatoid and/or poor-risk metastatic renal cell carcinoma. Cancer 121, 3435–43. doi:10.1002/cncr.29503

Mikeska, T., Craig, J.M., 2014. DNA methylation biomarkers: cancer and beyond. Genes (Basel). 5, 821–64. doi:10.3390/genes5030821

Milan, G., Romanello, V., Pescatore, F., Armani, A., Paik, J.-H., Frasson, L., Seydel, A., Zhao, J., Abraham, R., Goldberg, A.L., Blaauw, B., DePinho, R.A., Sandri, M., 2015. Regulation of autophagy and the ubiquitin-proteasome system by the FoxO transcriptional network during muscle atrophy. Nat. Commun. 6, 6670. doi:10.1038/ncomms7670

Millet, I., Doyon, F.C., Hoa, D., Thuret, R., Merigeaud, S., Serre, I., Taourel, P., 2011. Characterization of small solid renal lesions: can benign and malignant tumors be differentiated with CT? AJR. Am. J. Roentgenol. 197, 887–96. doi:10.2214/AJR.10.6276

Mizuno, K., Osada, H., Konishi, H., Tatematsu, Y., Yatabe, Y., Mitsudomi, T., Fujii, Y., Takahashi, T., 2002. Aberrant hypermethylation of the CHFR prophase checkpoint gene in human lung cancers. Oncogene 21, 2328–2333. doi:10.1038/sj.onc.1205402

Morris, M.R., Gentle, D., Abdulrahman, M., Maina, E.N., Gupta, K., Banks, R.E., Wiesener, M.S., Kishida, T., Yao, M., Teh, B., Latif, F., Maher, E.R., 2005. Tumor suppressor activity and epigenetic inactivation of hepatocyte growth factor activator inhibitor type 2/SPINT2 in papillary and clear cell renal cell carcinoma. Cancer Res. 65, 4598–606. doi:10.1158/0008-5472.CAN-04-3371

Morris, M.R., Maher, E.R., 2010. Epigenetics of renal cell carcinoma: the path towards new diagnostics and therapeutics. Genome Med. 2, 59. doi:10.1186/gm180

Morris, M.R., Ricketts, C., Gentle, D., Abdulrahman, M., Clarke, N., Brown, M., Kishida, T., Yao, M., Latif, F., Maher, E.R., 2010. Identification of candidate tumour suppressor genes frequently methylated in renal cell carcinoma. Oncogene 29, 2104–17. doi:10.1038/onc.2009.493

Morris, M.R., Ricketts, C.J., Gentle, D., McRonald, F., Carli, N., Khalili, H., Brown, M., Kishida, T., Yao, M., Banks, R.E., Clarke, N., Latif, F., Maher, E.R., 2011. Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. Oncogene 30, 1390–1401. doi:10.1038/onc.2010.525

Morris, T.J., Beck, S., 2015. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. Methods 72, 3–8. doi:10.1016/j.ymeth.2014.08.011

Morrissey, C., Martinez, A., Zatyka, M., Agathanggelou, A., Honorio, S., Astuti, D., Morgan, N. V, Moch, H., Richards, F.M., Kishida, T., Yao, M., Schraml, P., Latif, F., Maher, E.R., 2001. Epigenetic Inactivation of the RASSF1A 3p21 . 3 Tumor Suppressor Gene in Both Clear Cell and Papillary Renal Cell Carcinoma 1. Cancer Res. 61, 7277–7281.

Muller, P.A.J., Vousden, K.H., 2013. p53 mutations in cancer. Nat. Cell Biol. 15, 2–8. doi:10.1038/ncb2641

Murakami, T., Sano, F., Huang, Y., Komiya, A., Baba, M., Osada, Y., Nagashima, Y., Kondo, K., Nakaigawa, N., Miura, T., Kubota, Y., Yao, M., Kishida, T., 2007. Identification and

characterization of Birt-Hogg-Dubé associated renal carcinoma. J. Pathol. 211, 524–31. doi:10.1002/path.2139

Nagasaki, M., Yasuda, J., Katsuoka, F., Nariai, N., Kojima, K., Kawai, Y., Yamaguchi-Kabata, Y., Yokozawa, J., Danjoh, I., Saito, S., Sato, Y., Mimori, T., Tsuda, K., Saito, R., Pan, X., Nishikawa, S., Ito, S., Kuroki, Y., Tanabe, O., Fuse, N., Kuriyama, S., Kiyomoto, H., Hozawa, A., Minegishi, N., Douglas Engel, J., Kinoshita, K., Kure, S., Yaegashi, N., Yamamoto, M., 2015. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat. Commun. 6, 8018. doi:10.1038/ncomms9018

Nault, J.C., Mallet, M., Pilati, C., Calderaro, J., Bioulac-Sage, P., Laurent, C., Laurent, A., Cherqui, D., Balabaud, C., Zucman-Rossi, J., Zucman Rossi, J., 2013. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. Nat. Commun. 4, 2218. doi:10.1038/ncomms3218

Nazio, F., Strappazzon, F., Antonioli, M., Bielli, P., Cianfanelli, V., Bordi, M., Gretzmeier, C., Dengjel, J., Piacentini, M., Fimia, G.M., Cecconi, F., 2013. mTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6. Nat. Cell Biol. 15, 406–16. doi:10.1038/ncb2708

Neapolitan, R., Horvath, C.M., Jiang, X., 2015. Pan-cancer analysis of TCGA data reveals notable signaling pathways. BMC Cancer 15, 516. doi:10.1186/s12885-015-1484-6

Neuzillet, Y., Lechevallier, E., Andre, M., Daniel, L., Nahon, O., Coulange, C., 2005. Follow-up of renal oncocytoma diagnosed by percutaneous tumor biopsy. Urology 66, 1181–5. doi:10.1016/j.urology.2005.06.001

Ng, C.S., Wood, C.G., Silverman, P.M., Tannir, N.M., Tamboli, P., Sandler, C.M., 2008. Renal Cell Carcinoma: Diagnosis, Staging, and Surveillance. Am. J. Roentgenol. 191, 1220–1232. doi:10.2214/AJR.07.3568

Ng, K.L., Rajandram, R., Morais, C., Yap, N.Y., Samaratunga, H., Gobe, G.C., Wood, S.T., 2014. Differentiation of oncocytoma from chromophobe renal cell carcinoma (RCC): can novel molecular biomarkers help solve an old problem? J. Clin. Pathol. 67, 97–104. doi:10.1136/jclinpath-2013-201895

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., Shendure, J., Bamshad, M.J., 2010. Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet. 42, 30–35. doi:10.1038/ng.499

Nguyen, D.P., Vilaseca, A., Vertosick, E. a., Corradi, R.B., Touijer, K. a., Benfante, N.E., Sjoberg, D.D., Russo, P., 2015. Histologic subtype impacts cancer-specific survival in patients with sarcomatoid-variant renal cell carcinoma treated surgically. World J. Urol. doi:10.1007/s00345-015-1644-y

Nickerson, M.L., Warren, M.B., Toro, J.R., Matrosova, V., Glenn, G., Turner, M.L., Duray, P., Merino, M., Choyke, P., Pavlovich, C.P., Sharma, N., Walther, M., Munroe, D., Hill, R., Maher, E., Greenberg, C., Lerman, M.I., Linehan, W.M., Zbar, B., Schmidt, L.S., 2002. Mutations in a novel gene lead to kidney tumors, lung wall defects, and benign tumors of the hair follicle in patients with the Birt-Hogg-Dubé syndrome. Cancer Cell 2, 157–164. doi:10.1016/S1535-6108(02)00104-6

Nonaka, D., Chiriboga, L., Soslow, R.A., 2008. Expression of pax8 as a useful marker in distinguishing ovarian carcinomas from mammary carcinomas. Am. J. Surg. Pathol. 32, 1566–71. doi:10.1097/PAS.0b013e31816d71ad

Noushmehr, H., Weisenberger, D.J., Diefes, K., Phillips, H.S., Pujara, K., Berman, B.P., Pan, F., Pelloski, C.E., Sulman, E.P., Bhat, K.P., Verhaak, R.G.W., Hoadley, K.A., Hayes, D.N., Perou, C.M., Schmidt, H.K., Ding, L., Wilson, R.K., Van Den Berg, D., Shen, H., Bengtsson, H., Neuvial, P., Cope, L.M., Buckley, J., Herman, J.G., Baylin, S.B., Laird, P.W., Aldape, K., 2010. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell 17, 510–22. doi:10.1016/j.ccr.2010.03.017

Nowell, P.C., 1976. The clonal evolution of tumor cell populations. Science 194, 23–8.

Oda, H., Nakatsuru, Y., Ishikawa, T., 1995. Mutations of the p53 Gene and p53 Protein Overexpression Are Associated with Sarcomatoid Transformation in Renal Cell Carcinomas Mutations of the p53 Gene and p53 Protein Overexpression Are Associated with Sarcomatoid Transformation in Renal Cell Carcinom 658–662.

Office for National Statistics, 2015a. Cancer Registration Statistics, England, 2013 1–19.

Office for National Statistics, 2015b. Cancer Registration Statistics, England, 2013 1–19.

Office for National Statistics, Cancer Research UK Cancer Survival Group London School of Hygiene and Tropical Medicine, Exarchakou, A., Rachet, B., Nash, E., Bannister, N., Coleman, M.P., Rowlands, S., 2015. Cancer Survival in England: adults diagnosed in 2009 to 2013, followerd up to 2014 14.

Ogino, S., Kawasaki, T., Brahmandam, M., Cantor, M., Kirkner, G.J., Spiegelman, D., Makrigiorgos, G.M., Weisenberger, D.J., Laird, P.W., Loda, M., Fuchs, C.S., 2006. Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. J. Mol. Diagn. 8, 209–17. doi:10.2353/jmoldx.2006.050135

Olkhov-Mitsel, E., Zdravic, D., Kron, K., van der Kwast, T., Fleshner, N., Bapat, B., 2014. Novel multiplex MethyLight protocol for detection of DNA methylation in patient tissues and bodily fluids. Sci. Rep. 4, 4432. doi:10.1038/srep04432

Ooi, A., Dykema, K., Ansari, A., Petillo, D., Snider, J., Kahnoski, R., Anema, J., Craig, D., Carpten, J., Teh, B.T., Furge, K. a., 2013. CUL3 and NRF2 mutations confer an NRF2 activation phenotype in a sporadic form of papillary renal cell carcinoma. Cancer Res. 73, 2044–2051. doi:10.1158/0008-5472.CAN-12-3227

Pan, D., 2010. The hippo signaling pathway in development and cancer. Dev. Cell 19, 491–505. doi:10.1016/j.devcel.2010.09.011

Pancione, M., Remo, A., Zanella, C., Sabatino, L., Di Blasi, A., Laudanna, C., Astati, L., Rocco, M., Bifano, D., Piacentini, P., Pavan, L., Purgato, A., Greco, F., Talamini, A., Bonetti, A., Ceccarelli, M., Vendraminelli, R., Manfrin, E., Colantuoni, V., 2013. The chromatin remodelling component SMARCB1/INI1 influences the metastatic behavior of colorectal cancer through a gene signature mapping to chromosome 22. J. Transl. Med. 11, 297. doi:10.1186/1479-5876-11-297

Parkin, D.M., Boyd, L., 2011. 8. Cancers attributable to overweight and obesity in the UK in 2010. Br. J. Cancer 105 Suppl , S34–7. doi:10.1038/bjc.2011.481

Parkin, D.M., Boyd, L., Walker, L.C., 2011. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br. J. Cancer 105 Suppl , S77–81. doi:10.1038/bjc.2011.489

Patel, Z.H., Kottyan, L.C., Lazaro, S., Williams, M.S., Ledbetter, D.H., Tromp, H., Rupert, A., Kohram, M., Wagner, M., Husami, A., Qian, Y., Valencia, C.A., Zhang, K., Hostetter, M.K., Harley, J.B., Kaufman, K.M., 2014. The struggle to find reliable results in exome sequencing data: filtering out Mendelian errors. Front. Genet. 5, 16. doi:10.3389/fgene.2014.00016

Peedicayil, J., 2006. Epigenetic therapy--a new development in pharmacology. Indian J. Med. Res. 123, 17–24.

Peifer, M., Fernández-Cuesta, L., Sos, M.L., George, J., Seidel, D., Kasper, L.H., Plenker, D., Leenders, F., Sun, R., Zander, T., Menon, R., Koker, M., Dahmen, I., Müller, C., Di Cerbo, V., Schildhaus, H.-U., Altmüller, J., Baessmann, I., Becker, C., de Wilde, B., Vandesompele, J., Böhm, D., Ansén, S., Gabler, F., Wilkening, I., Heynck, S., Heuckmann, J.M., Lu, X., Carter, S.L., Cibulskis, K., Banerji, S., Getz, G., Park, K.-S., Rauh, D., Grütter, C., Fischer, M., Pasqualucci, L., Wright, G., Wainer, Z., Russell, P., Petersen, I., Chen, Y., Stoelben, E., Ludwig, C., Schnabel, P., Hoffmann, H., Muley, T., Brockmann, M., Engel-Riedel, W., Muscarella, L. a, Fazio, V.M., Groen, H., Timens, W.,

Sietsma, H., Thunnissen, E., Smit, E., Heideman, D.A.M., Snijders, P.J.F., Cappuzzo, F., Ligorio, C., Damiani, S., Field, J., Solberg, S., Brustugun, O.T., Lund-Iversen, M., Sänger, J., Clement, J.H., Soltermann, A., Moch, H., Weder, W., Solomon, B., Soria, J.-C., Validire, P., Besse, B., Brambilla, E., Brambilla, C., Lantuejoul, S., Lorimier, P., Schneider, P.M., Hallek, M., Pao, W., Meyerson, M., Sage, J., Shendure, J., Schneider, R., Büttner, R., Wolf, J., Nürnberg, P., Perner, S., Heukamp, L.C., Brindle, P.K., Haas, S., Thomas, R.K., 2012. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat. Genet. 44, 1104–1110. doi:10.1038/ng.2396

Peña-Llopis, S., Vega-Rubín-de-Celis, S., Liao, A., Leng, N., Pavía-Jiménez, A., Wang, S., Yamasaki, T., Zhrebker, L., Sivanand, S., Spence, P., Kinch, L., Hambuch, T., Jain, S., Lotan, Y., Margulis, V., Sagalowsky, A.I., Summerour, P.B., Kabbani, W., Wong, S.W.W., Grishin, N., Laurent, M., Xie, X.-J., Haudenschild, C.D., Ross, M.T., Bentley, D.R., Kapur, P., Brugarolas, J., 2012. BAP1 loss defines a new class of renal cell carcinoma. Nat. Genet. 44, 751–9. doi:10.1038/ng.2323

Peng, H., Liu, H., Zhao, S., Wu, J., Fan, J., Liao, J., 2013. Silencing of RASSF3 by DNA hypermethylation is associated with tumorigenesis in somatotroph adenomas. PLoS One 8, e59024. doi:10.1371/journal.pone.0059024

Peters, T.L., Kumar, V., Polikepahad, S., Lin, F.Y., Sarabia, S.F., Liang, Y., Wang, W.-L., Lazar, A.J., Doddapaneni, H., Chao, H., Muzny, D.M., Wheeler, D.A., Okcu, M.F., Plon, S.E., Hicks, M.J., López-Terrada, D., Parsons, D.W., Roy, A., 2015. BCOR-CCNB3 fusions are frequent in undifferentiated sarcomas of male children. Mod. Pathol. 28, 575–86. doi:10.1038/modpathol.2014.139

Petherick, K.J., Conway, O.J.L., Mpamhanga, C., Osborne, S.A., Kamal, A., Saxty, B., Ganley, I.G., 2015. Pharmacological Inhibition of ULK1 Blocks mTOR-Dependent Autophagy. J. Biol. Chem. 290, 11376–83. doi:10.1074/jbc.C114.627778

Pfeifer, G.P., Dammann, R., Tommasi, S., 2010. Quick guide RASSF proteins. Curr. Biol. 20, 344–345.

Pires-Luís, A.S., Vieira-Coimbra, M., Vieira, F.Q., Costa-Pinheiro, P., Silva-Santos, R., Dias, P.C., Antunes, L., Lobo, F., Oliveira, J., Gonçalves, C.S., Costa, B.M., Henrique, R., Jerónimo, C., 2015. Expression of histone methyltransferases as novel biomarkers for renal cell tumor diagnosis and prognostication. Epigenetics 2294. doi:10.1080/15592294.2015.1103578

Pon, J.R., Marra, M.A., 2015. Driver and passenger mutations in cancer. Annu. Rev. Pathol. 10, 25–50. doi:10.1146/annurev-pathol-012414-040312

Pylayeva-Gupta, Y., Grabocka, E., Bar-Sagi, D., 2011. RAS oncogenes: weaving a tumorigenic web. Nat. Rev. Cancer 11, 761–74. doi:10.1038/nrc3106

Quaresma, M., Coleman, M.P., Rachet, B., 2015. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971–2011: a population-based study. Lancet 385, 1206–1218. doi:10.1016/S0140-6736(14)61396-9

Rabbani, B., Mahdieh, N., Hosomichi, K., Nakaoka, H., Inoue, I., 2012. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. J. Hum. Genet. 57, 621–632. doi:10.1038/jhg.2012.91

Rabbani, B., Tekin, M., Mahdieh, N., 2014a. The promise of whole-exome sequencing in medical genetics. J. Hum. Genet. 59, 5–15. doi:10.1038/jhg.2013.114

Rabbani, B., Tekin, M., Mahdieh, N., 2014b. The promise of whole-exome sequencing in medical genetics. J. Hum. Genet. 59, 5–15. doi:10.1038/jhg.2013.114

Rajalingam, K., Schreck, R., Rapp, U.R., Albert, Š., 2007. Ras oncogenes and their downstream targets. Biochim. Biophys. Acta - Mol. Cell Res. 1773, 1177–1195. doi:10.1016/j.bbamcr.2007.01.012

Ramakrishnan, S., Pili, R., 2013. Histone deacetylase inhibitors and epigenetic modifications

as a novel strategy in renal cell carcinoma. Cancer J. 19, 333–40. doi:10.1097/PPO.0b013e3182a09e07

Raphael, B.J., Dobson, J.R., Oesper, L., Vandin, F., 2014. Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine. Genome Med. 6, 5. doi:10.1186/gm524

Rathmell, K.W., Chen, F., Creighton, C.J., 2015. Genomics of chromophobe renal cell carcinoma: implications from a rare tumor for pan-cancer studies. Oncoscience 2, 81–90.

Recino, A., Sherwood, V., Flaxman, A., Cooper, W.N., Latif, F., Ward, A., Chalmers, A.D., 2010. Human RASSF7 regulates the microtubule cytoskeleton and is required for spindle formation, Aurora B activation and chromosomal congression during mitosis. Biochem. J. 430, 207–213. doi:10.1042/BJ20100883

Reese, M.G., Eeckman, F.H., Kulp, D., Haussler, D., 1997. Improved splice site detection in Genie. J. Comput. Biol. 4, 311–323. doi:10.1089/cmb.1997.4.311

Reis-Filho, J.S., 2009. Next-generation sequencing. Breast Cancer Res. 11 Suppl 3, S12. doi:10.1186/bcr2431

Renal cancer | Guidance and guideline topic | NICE, 2016.

Richter, A.M., Haag, T., Walesch, S., Herrmann-Trost, P., Marsch, W.C., Kutzner, H., Helmbold, P., Dammann, R.H., 2013. Aberrant Promoter Hypermethylation of RASSF Family Members in Merkel Cell Carcinoma. Cancers (Basel). 5, 1566–76. doi:10.3390/cancers5041566

Richter, A.M., Pfeifer, G.P., Dammann, R.H., 2009. The RASSF proteins in cancer; from epigenetic silencing to functional characterization. Biochim. Biophys. Acta - Rev. Cancer 1796, 114–128.

Richter, A.M., Walesch, S.K., Würl, P., Taubert, H., Dammann, R.H., 2012. The tumor suppressor RASSF10 is upregulated upon contact inhibition and frequently epigenetically silenced in cancer. Oncogenesis 1, e18. doi:10.1038/oncsis.2012.18

Richter, A.M., Zimmermann, T., Haag, T., Walesch, S.K., Dammann, R.H., 2015. Promoter methylation status of Ras-association domain family members in pheochromocytoma. Front. Endocrinol. (Lausanne). 6, 21. doi:10.3389/fendo.2015.00021

Ricketts, C.J., Hill, V.K., Linehan, W.M., 2014. Tumor-specific hypermethylation of epigenetic biomarkers, including SFRP1, predicts for poorer survival in patients from the TCGA Kidney Renal Clear Cell Carcinoma (KIRC) project. PLoS One 9, e85621. doi:10.1371/journal.pone.0085621

Ricketts, C.J., Morris, M.R., Gentle, D., Brown, M., Wake, N., Woodward, E.R., Clarke, N., Latif, F., Maher, E.R., 2012. Genome-wide CpG island methylation analysis implicates novel genes in the pathogenesis of renal cell carcinoma. Epigenetics 7, 278–90. doi:10.4161/epi.7.3.19103

Ricketts, C.J., Morris, M.R., Gentle, D., Shuib, S., Brown, M., Clarke, N., Wei, W., Nathan, P., Latif, F., Maher, E.R., 2013. Methylation profiling and evaluation of demethylating therapy in renal cell carcinoma. Clin. Epigenetics 5, 16. doi:10.1186/1868-7083-5-16

Rivlin, N., Brosh, R., Oren, M., Rotter, V., 2011. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. Genes Cancer 2, 466–74. doi:10.1177/1947601911408889

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–6. doi:10.1038/nbt.1754

Rodríguez-Paredes, M., Esteller, M., 2011. Cancer epigenetics reaches mainstream oncology. Nat. Med. 17, 330–339. doi:10.1038/nm.2305

Rogenhofer, S., Kahl, P., Holzapfel, S., von Ruecker, A., Mueller, S.C., Ellinger, J., 2012a. Decreased Levels of Histone H3K9me1 Indicate Poor Prognosis in Patients with Renal Cell Carcinoma. Anticancer Res. 32, 879–886.

Rogenhofer, S., Kahl, P., Mertens, C., Hauser, S., Hartmann, W., Büttner, R., Müller, S., Von Ruecker, A., Ellinger, J., 2012b. Global histone H3 lysine 27 (H3K27) methylation levels and their prognostic relevance in renal cell carcinoma. BJU Int. 109, 459–465. doi:10.1111/j.1464-410X.2011.10278.x

Romano, D., Matallanas, D., Frederick, D.T., Flaherty, K.T., Kolch, W., 2014. One Hippo and many masters: differential regulation of the Hippo pathway in cancer: Figure 1. Biochem. Soc. Trans. 42, 816–821. doi:10.1042/BST20140030

Romano, D., Matallanas, D., Weitsman, G., Preisinger, C., Ng, T., Kolch, W., 2010. Proapoptotic kinase MST2 coordinates signaling crosstalk between RASSF1A, Raf-1, and Akt. Cancer Res. 70, 1195–203. doi:10.1158/0008-5472.CAN-09-3147

Rose, N.R., Klose, R.J., 2014. Understanding the relationship between DNA methylation and histone lysine methylation. Biochim. Biophys. Acta 1839, 1362–1372. doi:10.1016/j.bbagrm.2014.02.007

Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G., Lee, B.T., Barber, G.P., Harte, R.A., Diekhans, M., Long, J.C., Wilder, S.P., Zweig, A.S., Karolchik, D., Kuhn, R.M., Haussler, D., Kent, W.J., 2013. ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res. 41, D56–63. doi:10.1093/nar/gks1172

Rosner, M., Schipany, K., Hengstschläger, M., 2013. Merging high-quality biochemical fractionation with a refined flow cytometry approach to monitor nucleocytoplasmic protein expression throughout the unperturbed mammalian cell cycle. Nat. Protoc. 8, 602–626. doi:10.1038/nprot.2013.011

Roy, D.M., Walsh, L.A., Chan, T.A., 2014. Driver mutations of cancer epigenomes. Protein Cell 5, 265–96. doi:10.1007/s13238-014-0031-6

Ruddon, R.W., 2007. Cancer Biology. Oxford University Press, USA.

Sajadian, S.O., Ehnert, S., Vakilian, H., Koutsouraki, E., Damm, G., Seehofer, D., Thasler, W., Dooley, S., Baharvand, H., Sipos, B., Nussler, A.K., 2015. Induction of active demethylation and 5hmC formation by 5-azacytidine is TET2 dependent and suggests new treatment strategies against hepatocellular carcinoma. Clin. Epigenetics 7, 98. doi:10.1186/s13148-015-0133-x

Saldanha, A.J., 2004. Java Treeview - Extensible visualization of microarray data. Bioinformatics 20, 3246–3248. doi:10.1093/bioinformatics/bth349

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., Esteller, M., 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6, 692–702.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–7.

Sato, F., Tsuchiya, S., Meltzer, S.J., Shimizu, K., 2011. MicroRNAs and epigenetics. FEBS J. 278, 1598–609. doi:10.1111/j.1742-4658.2011.08089.x

Sato, Y., Yoshizato, T., Shiraishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H., Nagata, Y., Yoshida, K., Kon, A., Suzuki, Y., Chiba, K., Tanaka, H., Niida, A., Fujimoto, A., Tsunoda, T., Morikawa, T., Maeda, D., Kume, H., Sugano, S., Fukayama, M., Aburatani, H., Sanada, M., Miyano, S., Homma, Y., Ogawa, S., 2013. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat. Genet. 45, 860–7. doi:10.1038/ng.2699

Sawan, C., Herceg, Z., 2010. Histone modifications and cancer. Adv. Genet. 70, 57–85. doi:10.1016/B978-0-12-380866-0.60003-4

Schagdarsurengin, U., Richter, A.M., Hornung, J., Lange, C., Steinmann, K., Dammann, R.H., 2010. Frequent epigenetic inactivation of RASSF2 in thyroid cancer and functional consequences. Mol. Cancer 9, 264. doi:10.1186/1476-4598-9-264

Schütte, U., Bisht, S., Heukamp, L.C., Kebschull, M., Florin, A., Haarmann, J., Hoffmann, P., Bendas, G., Buettner, R., Brossart, P., Feldmann, G., 2014. Hippo signaling mediates proliferation, invasiveness, and metastatic potential of clear cell renal cell carcinoma. Transl. Oncol. 7, 309–21. doi:10.1016/j.tranon.2014.02.005

Scott, S.A., 2011. Personalizing medicine with clinical pharmacogenetics. Genet. Med. 13, 987–95. doi:10.1097/GIM.0b013e318238b38c

Seidel, C., Schagdarsurengin, U., Blümke, K., Würl, P., Pfeifer, G.P., Hauptmann, S., Taubert, H., Dammann, R., 2007. Frequent hypermethylation of MST1 and MST2 in soft tissue sarcoma. Mol. Carcinog. 46, 865–71. doi:10.1002/mc.20317

Selamat, S.A., Galler, J.S., Joshi, A.D., Fyfe, M.N., Campan, M., Siegmund, K.D., Kerr, K.M., Laird-Offringa, I.A., 2011. DNA methylation changes in atypical adenomatous hyperplasia, adenocarcinoma in situ, and lung adenocarcinoma. PLoS One 6, e21443. doi:10.1371/journal.pone.0021443

Sévenet, N., Lellouch-Tubiana, A., Schofield, D., Hoang-Xuan, K., Gessler, M., Birnbaum, D., Jeanpierre, C., Jouvet, A., Delattre, O., 1999. Spectrum of hSNF5/INI1 somatic mutations in human cancer and genotype-phenotype correlations. Hum. Mol. Genet. 8, 2359–68.

Sharma, S., Kelly, T.K., Jones, P.A., 2010. Epigenetics in cancer. Carcinogenesis 31, 27–36. doi:10.1093/carcin/bgp220

Shinawi, T., Hill, V.K., Krex, D., Schackert, G., Gentle, D., Morris, M.R., Wei, W., Cruickshank, G., Maher, E.R., Latif, F., 2013. DNA methylation profiles of long- and short-term glioblastoma survivors. Epigenetics 8, 149–156. doi:10.4161/epi.23398

Shivakumar, L., Minna, J., Sakamaki, T., Pestell, R., White, M.A., 2002. The RASSF1A tumor suppressor blocks cell cycle progression and inhibits cyclin D1 accumulation. Mol. Cell. Biol. 22, 4309–18.

Shuch, B., Amin, A., Armstrong, A.J., Eble, J.N., Ficarra, V., Lopez-Beltran, A., Martignoni, G., Rini, B.I., Kutikov, A., 2015. Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities from biologic complexity. Eur. Urol. 67, 85–97. doi:10.1016/j.eururo.2014.04.029

Shuch, B., Bratslavsky, G., Linehan, W.M., Srinivasan, R., 2012a. Sarcomatoid renal cell carcinoma: a comprehensive review of the biology and current treatment strategies. Oncologist 17, 46–54. doi:10.1634/theoncologist.2011-0227

Shuch, B., Vourganti, S., Friend, J.C., Zehngebot, L.M., Linehan, W.M., Srinivasan, R., 2012b. Targeting the mTOR pathway in Chromophobe Kidney Cancer. J. Cancer 3, 152–7. doi:10.7150/jca.4378

Simon, R., Roychowdhury, S., 2013. Implementing personalized cancer genomics in clinical trials. Nat. Rev. Drug Discov. 12, 358–69. doi:10.1038/nrd3979

Sims III, R.J., Reinberg, D., 2009. Processing the H3K36me3 signature. Nat. Genet. 41, 270–271. doi:10.1038/ng0309-270

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Publ. Gr. 15. doi:10.1038/nrg3642

Siva, N., 2015. UK gears up to decode 100,000 genomes from NHS patients. Lancet (London, England) 385, 103–4. doi:10.1016/S0140-6736(14)62453-3

Slater, A.A., Alokail, M., Gentle, D., Yao, M., Kovacs, G., Maher, E.R., Latif, F., 2013. DNA methylation profiling distinguishes histological subtypes of renal cell carcinoma. Epigenetics 8, 252–67. doi:10.4161/epi.23817

Sottocornola, R., Royer, C., Vives, V., Tordella, L., Zhong, S., Wang, Y., Ratnayaka, I., Shipman, M., Cheung, A., Gaston-Massuet, C., Ferretti, P., Molnár, Z., Lu, X., 2010. ASPP2 binds Par-3 and controls the polarity and proliferation of neural progenitors during CNS development. Dev. Cell 19, 126–37. doi:10.1016/j.devcel.2010.06.003

Stebbing, J., Zhang, H., Xu, Y., Lit, L.C., Green, A.R., Grothey, A., Lombardo, Y., Periyasamy, M., Blighe, K., Zhang, W., Shaw, J.A., Ellis, I.O., Lenz, H.J., Giamas, G., 2015. KSR1 regulates BRCA1 degradation and inhibits breast cancer growth. Oncogene 34, 2103–14. doi:10.1038/onc.2014.129

Stec, R., Grala, B., Maczewski, M., Bodnar, L., Szczylik, C., 2009. Chromophobe renal cell cancer--review of the literature and potential methods of treating metastatic disease. J. Exp. Clin. Cancer Res. 28, 134. doi:10.1186/1756-9966-28-134

Stefansson, O.A., Moran, S., Gomez, A., Sayols, S., et al. 2015. A DNA methylation-based definition of biologically distinct breast cancer subtypes. Mol. Oncol. 9, 555–68. doi:10.1016/j.molonc.2014.10.012

Stéhelin, D., 1995. Oncogenes and cancer. Science 267, 1408–1409. doi:10.1126/science.7878455

Stewart, B., Wild, C.P., 2014. World Cancer Report 2014 (ePUB) - WHO - OMS -, eBook. International Agency for Reseacg on Cancer World Health organization.

Strimbu, K., Tavel, J.A., 2010. What are biomarkers? Curr. Opin. HIV AIDS 5, 463–6. doi:10.1097/COH.0b013e32833ed177

Su, Z., Chen, D., Zhang, E., Li, Y., Yu, Z., Shi, M., Jiang, Z., Ni, L., Yang, S., Gui, Y., Ye, J., Lai, Y., 2015. MicroRNA-509-3p inhibits cancer cell proliferation and migration by targeting the mitogen-activated protein kinase kinase kinase 8 oncogene in renal cell carcinoma. Mol. Med. Rep. 12, 1535–43. doi:10.3892/mmr.2015.3498

Subramaniam, D., Thombre, R., Dhar, A., Anant, S., 2014. DNA methyltransferases: a novel target for prevention and therapy. Front. Oncol. 4, 80. doi:10.3389/fonc.2014.00080

Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., Lehner, B., 2014. Synonymous mutations frequently act as driver mutations in human cancers. Cell 156, 1324–35. doi:10.1016/j.cell.2014.01.051

Suzuki, T., Tsurusaki, Y., Nakashima, M., Miyake, N., Saitsu, H., Takeda, S., Matsumoto, N., 2014. Precise detection of chromosomal translocation or inversion breakpoints by whole-genome sequencing. J. Hum. Genet. 59, 649–54. doi:10.1038/jhg.2014.88

Swerdlow, A.J., Peto, R., Doll, R.S., 2010. Epidemiology of cancer - Oxford Textbook of Medicine, 5th ed. Oxford University Press. doi:10.1093/med/9780199204854.003.0601

Tabassum, D.P., Polyak, K., 2015. Tumorigenesis: it takes a village. Nat. Rev. Cancer 15, 473–483. doi:10.1038/nrc3971

Tamborero, D., Abel, G., Lopez-Bigas, N., 2014. Identification of oncogenic driver mutations. Exp. Med. 32, 1–8.

Tamura, M., Sasaki, Y., Koyama, R., Takeda, K., Idogawa, M., Tokino, T., 2014. Forkhead transcription factor FOXF1 is a novel target gene of the p53 family and regulates cancer cell migration and invasiveness. Oncogene 33, 4837–46. doi:10.1038/onc.2013.427

Tan, M.-H., Wong, C.F., Tan, H.L., Yang, X.J., Ditlev, J., Matsuda, D., Khoo, S.K., Sugimura, J., Fujioka, T., Furge, K. a, Kort, E., Giraud, S., Ferlicot, S., Vielh, P., Amsellem-Ouazana, D., Debré, B., Flam, T., Thiounn, N., Zerbib, M., Benoît, G., Droupy, S., Molinié, V., Vieillefond, A., Tan, P.H., Richard, S., Teh, B.T., 2010. Genomic expression and single-nucleotide polymorphism profiling discriminates chromophobe renal cell carcinoma and oncocytoma. BMC Cancer 10, 196. doi:10.1186/1471-2407-10-196

Tawamie, H., Wohlleber, E., Uebe, S., Schmäl, C., Nöthen, M.M., Abou Jamra, R., 2015. Recurrent null mutation in SPG20 leads to Troyer syndrome. Mol. Cell. Probes 4–7. doi:10.1016/j.mcp.2015.05.006

TCGA, 2016. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N. Engl. J. Med. 374, 135–145. doi:10.1056/NEJMoa1505917

TCGA, 2014. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature 507, 315–322. doi:10.1038/nature12965

TCGA, T.C.G.A.R.N., 2013. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499, 43–9. doi:10.1038/nature12222

TCGA, T.C.G.A.R.N., 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455, 1061–8. doi:10.1038/nature07385

Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., Beck, S., 2013. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 29, 189–96. doi:10.1093/bioinformatics/bts680

Thermo Fischer Scientific, 2015. Ion AmpliSeq™ Cancer Hotspot Panel v2.

Thermo Fischer Scientific Inc., 2008. NanoDrop 1000 Spectrophotometer V3. 7 User's Manual. Thermo Fish. Sci. 5–1 to 5–4.

Thomas, R.K., Baker, A.C., Debiasi, R.M., Winckler, W., Laframboise, T., Lin, W.M., Wang, M., Feng, W., Zander, T., MacConaill, L., Macconnaill, L.E., Lee, J.C., Nicoletti, R., Hatton, C., Goyette, M., Girard, L., Majmudar, K., Ziaugra, L., Wong, K.-K., Gabriel, S., Beroukhim, R., Peyton, M., Barretina, J., Dutt, A., Emery, C., Greulich, H., Shah, K., Sasaki, H., Gazdar, A., Minna, J., Armstrong, S.A., Mellinghoff, I.K., Hodi, F.S., Dranoff, G., Mischel, P.S., Cloughesy, T.F., Nelson, S.F., Liau, L.M., Mertz, K., Rubin, M.A., Moch, H., Loda, M., Catalona, W., Fletcher, J., Signoretti, S., Kaye, F., Anderson, K.C., Demetri, G.D., Dummer, R., Wagner, S., Herlyn, M., Sellers, W.R., Meyerson, M., Garraway, L.A., 2007. High-throughput oncogene mutation profiling in human cancer. Nat. Genet. 39, 347–51. doi:10.1038/ng1975

Tian, W., Li, W., Chen, Y., Yan, Z., Huang, X., Zhuang, H., Zhong, W., Chen, Y., Wu, W., Lin, C., Chen, H., Hou, X., Zhang, L., Sui, S., Zhao, B., Hu, Z., Li, L., Feng, D., 2015. Phosphorylation of ULK1 by AMPK regulates translocation of ULK1 to mitochondria and mitophagy. FEBS Lett. doi:10.1016/j.febslet.2015.05.020

Tost, J., Gut, I.G., 2007. DNA methylation analysis by pyrosequencing. Nat. Protoc. 2, 2265–75. doi:10.1038/nprot.2007.314

Totoki, Y., Tatsuno, K., Covington, K.R., Ueda, H., Creighton, C.J., Kato, M., Tsuji, S., Donehower, L.A., Slagle, B.L., Nakamura, H., Yamamoto, S., Shinbrot, E., Hama, N., Lehmkuhl, M., Hosoda, F., Arai, Y., Walker, K., Dahdouli, M., Gotoh, K., Nagae, G., Gingras, M.-C., Muzny, D.M., Ojima, H., Shimada, K., Midorikawa, Y., Goss, J.A., Cotton, R., Hayashi, A., Shibahara, J., Ishikawa, S., Guiteau, J., Tanaka, M., Urushidate, T., Ohashi, S., Okada, N., Doddapaneni, H., Wang, M., Zhu, Y., Dinh, H., Okusaka, T., Kokudo, N., Kosuge, T., Takayama, T., Fukayama, M., Gibbs, R.A., Wheeler, D.A., Aburatani, H., Shibata, T., 2014. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat. Genet. 46, 1267–73. doi:10.1038/ng.3126

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., Issa, J.P., 1999. CpG island methylator phenotype in colorectal cancer. Proc. Natl. Acad. Sci. U. S. A. 96, 8681–6.

Trabelsi, S., Mama, N., Ladib, M., Karmeni, N., Haddaji Mastouri, M., Chourabi, M., Mokni, M., Tlili, K., Krifa, H., Yacoubi, M.T., Saad, A., H'mida Ben Brahim, D., 2015. MGMT methylation assessment in glioblastoma: MS-MLPA versus human methylation 450K beadchip array and immunohistochemistry. Clin. Transl. Oncol. doi:10.1007/s12094-015-1381-0

Trinh, B.N., Long, T.I., Laird, P.W., 2001. DNA methylation analysis by MethyLight technology. Methods 25, 456–62. doi:10.1006/meth.2001.1268

Tuupanen, S., Turunen, M., Lehtonen, R., Hallikas, O., Vanharanta, S., Kivioja, T., Björklund, M., Wei, G., Yan, J., Niittymäki, I., Mecklin, J.-P., Järvinen, H., Ristimäki, A., Di-Bernardo, M., East, P., Carvajal-Carmona, L., Houlston, R.S., Tomlinson, I., Palin, K., Ukkonen, E., Karhu, A., Taipale, J., Aaltonen, L.A., 2009. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat. Genet. 41, 885–90. doi:10.1038/ng.406

Ueno-Yokohata, H., Okita, H., Nakasato, K., Akimoto, S., Hata, J., Koshinaga, T., Fukuzawa, M., Kiyokawa, N., 2015. Consistent in-frame internal tandem duplications of BCOR characterize clear cell sarcoma of the kidney. Nat. Genet. 2–5. doi:10.1038/ng.3338

Underhill-Day, N., Hill, V., Latif, F., 2011a. N-terminal RASSF family: RASSF7-RASSF10. Epigenetics 6, 284–92.

Underhill-Day, N., Hill, V., Latif, F., 2011b. N-terminal RASSF family: RASSF7-RASSF10. Epigenetics 6, 284–92.

van Haaften, G., Dalgliesh, G.L., Davies, H., Chen, L., Bignell, G., Greenman, C., Edkins, S., Hardy, C., O'Meara, S., Teague, J., Butler, A., Hinton, J., Latimer, C., Andrews, J., Barthorpe, S., Beare, D., Buck, G., Campbell, P.J., Cole, J., Forbes, S., Jia, M., Jones, D., Kok, C.Y., Leroy, C., Lin, M.-L., McBride, D.J., Maddison, M., Maquire, S., McLay, K., Menzies, A., Mironenko, T., Mulderrig, L., Mudie, L., Pleasance, E., Shepherd, R., Smith, R., Stebbings, L., Stephens, P., Tang, G., Tarpey, P.S., Turner, R., Turrell, K., Varian, J., West, S., Widaa, S., Wray, P., Collins, V.P., Ichimura, K., Law, S., Wong, J., Yuen, S.T., Leung, S.Y., Tonon, G., DePinho, R. a, Tai, Y.-T., Anderson, K.C., Kahnoski, R.J., Massie, A., Khoo, S.K., Teh, B.T., Stratton, M.R., Futreal, P.A., 2009. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat. Genet. 41, 521–523. doi:10.1038/ng.349

van Veldhoven, K., Polidoro, S., Baglietto, L., Severi, G., Sacerdote, C., Panico, S., Mattiello, A., Palli, D., Masala, G., Krogh, V., Agnoli, C., Tumino, R., Frasca, G., Flower, K., Curry, E., Orr, N., Tomczyk, K., Jones, M.E., Ashworth, A., Swerdlow, A., Chadeau-Hyam, M., Lund, E., Garcia-Closas, M., Sandanger, T.M., Flanagan, J.M., Vineis, P., 2015. Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. - PubMed - NCBI [WWW Document]. Clin. Epigenetics. doi:10.1186/s13148-015-0104-2

Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C.K., Stephens, P., Davies, H., Jones, D., Lin, M.-L., Teague, J., Bignell, G., Butler, A., Cho, J., Dalgliesh, G.L., Galappaththige, D., Greenman, C., Hardy, C., Jia, M., Latimer, C., Lau, K.W., Marshall, J., McLaren, S., Menzies, A., Mudie, L., Stebbings, L., Largaespada, D.A., Wessels, L.F.A., Richard, S., Kahnoski, R.J., Anema, J., Tuveson, D.A., Perez-Mancera, P.A., Mustonen, V., Fischer, A., Adams, D.J., Rust, A., Chan-on, W., Subimerb, C., Dykema, K., Furge, K., Campbell, P.J., Teh, B.T., Stratton, M.R., Futreal, P.A., 2011. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 469, 539–42. doi:10.1038/nature09639

Venkatesh, S., Workman, J.L., 2015. Histone exchange, chromatin structure and the regulation of transcription. Nat. Rev. Mol. Cell Biol. 16, 178–189. doi:10.1038/nrm3941

Vera-Badillo, F.E., Conde, E., Duran, I., 2012. Chromophobe renal cell carcinoma: A review of an uncommon entity. Int. J. Urol. 19, 894–900. doi:10.1111/j.1442-2042.2012.03079.x

Versteege, I., Sévenet, N., Lange, J., Rousseau-Merck, M.F., Ambros, P., Handgretinger, R., Aurias, A., Delattre, O., 1998. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. Nature 394, 203–6. doi:10.1038/28212

Vidarsson, H., Westergren, R., Heglind, M., Blomqvist, S.R., Breton, S., Enerbäck, S., 2009. The forkhead transcription factor Foxi1 is a master regulator of vacuolar H-ATPase proton pump subunits in the inner ear, kidney and epididymis. PLoS One 4, e4471. doi:10.1371/journal.pone.0004471

Vinagre, J., Almeida, A., Pópulo, H., Batista, R., Lyra, J., Pinto, V., Coelho, R., Celestino, R., Prazeres, H., Lima, L., Melo, M., da Rocha, A.G., Preto, A., Castro, P., Castro, L., Pardal, F., Lopes, J.M., Santos, L.L., Reis, R.M., Cameselle-Teijeiro, J., Sobrinho-Simões, M., Lima, J., Máximo, V., Soares, P., 2013. Frequency of TERT promoter mutations in human cancers. Nat. Commun. 4, 2185. doi:10.1038/ncomms3185

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Kinzler, K.W., 2013. Cancer genome landscapes. Science 339, 1546–58. doi:10.1126/science.1235122

Waldmann, T., Schneider, R., 2013. Targeting histone modifications--epigenetics in cancer.

Curr. Opin. Cell Biol. 25, 184–9. doi:10.1016/j.ceb.2013.01.001

Waldrop, M.M., 2016. The chips are down for Moore's law. Nat. News 530, 144. doi:10.1038/530144a

Wang, J., Andrici, J., Sioson, L., Clarkson, A., Sheen, A., Farzin, M., Toon, C.W., Turchini, J., Gill, A.J., 2016. Loss of INI1 expression in colorectal carcinoma is associated with high tumor grade, poor survival, BRAFV600E mutation and mismatch repair deficiency. Hum. Pathol. doi:10.1016/j.humpath.2016.04.018

Wang, K., Liu, T., Liu, L., Liu, J., Liu, C., Wang, C., Ge, N., Ren, H., Yan, K., Hu, S., Björkholm, M., Fan, Y., Xu, D., 2014. TERT promoter mutations in renal cell carcinomas and upper tract urothelial carcinomas. Oncotarget.

Wang, S.-S., Gu, Y.-F., Wolff, N., Stefanius, K., Christie, A., Dey, A., Hammer, R.E., Xie, X.-J., Rakheja, D., Pedrosa, I., Carroll, T., McKay, R.M., Kapur, P., Brugarolas, J., 2014. Bap1 is essential for kidney function and cooperates with Vhl in renal tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. 111, 16538–43. doi:10.1073/pnas.1414789111

Wang, Y., Godin-Heymann, N., Dan Wang, X., Bergamaschi, D., Llanos, S., Lu, X., 2013. ASPP1 and ASPP2 bind active RAS, potentiate RAS signalling and enhance p53 activity in cancer cells. Cell Death Differ. 20, 525–34. doi:10.1038/cdd.2013.3

Wang, Z., Liu, Y., Takahashi, M., Van Hook, K., Kampa-Schittenhelm, K.M., Sheppard, B.C., Sears, R.C., Stork, P.J.S., Lopez, C.D., 2013. N terminus of ASPP2 binds to Ras and enhances Ras/Raf/MEK/ERK activation to promote oncogene-induced senescence. Proc. Natl. Acad. Sci. U. S. A. 110, 312–7. doi:10.1073/pnas.1201514110

Watson, I.R., Takahashi, K., Futreal, P.A., Chin, L., 2013. Emerging patterns of somatic mutations in cancer. Nat. Rev. Genet. 14, 703–18. doi:10.1038/nrg3539

Watt, P.M., Kumar, R., Kees, U.R., 2000. Promoter demethylation accompanies reactivation of the HOX11 proto-oncogene in leukemia. Genes. Chromosomes Cancer 29, 371–7.

Wei, Z., Chen, X., Chen, J., Wang, W., Xu, X., Cai, Q., 2013. RASSF10 is epigenetically silenced and functions as a tumor suppressor in gastric cancer. Biochem. Biophys. Res. Commun. 432, 632–637. doi:10.1016/j.bbrc.2013.02.033

Weinberg, R., 2013. The Biology of Cancer, Second Edition. Garland Science.

Weinreb, I., Piscuoglio, S., Martelotto, L.G., Waggott, D., Ng, C.K.Y., Perez-Ordonez, B., Harding, N.J., Alfaro, J., Chu, K.C., Viale, A., Fusco, N., da Cruz Paula, A., Marchio, C., Sakr, R.A., Lim, R., Thompson, L.D.R., Chiosea, S.I., Seethala, R.R., Skalova, A., Stelow, E.B., Fonseca, I., Assaad, A., How, C., Wang, J., de Borja, R., Chan-Seng-Yue, M., Howlett, C.J., Nichols, A.C., Wen, Y.H., Katabi, N., Buchner, N., Mullen, L., Kislinger, T., Wouters, B.G., Liu, F.-F., Norton, L., McPherson, J.D., Rubin, B.P., Clarke, B.A., Weigelt, B., Boutros, P.C., Reis-Filho, J.S., 2014. Hotspot activating PRKD1 somatic mutations in polymorphous low-grade adenocarcinomas of the salivary glands. Nat. Genet. 46, 1166–9. doi:10.1038/ng.3096

Weisenberger, D.J., Van, D., Berg, D., Pan, F., Berman, B.P., Laird, P.W., 2008. Comprehensive DNA Methylation Analysis on the Illumina ® Infinium ® Assay Platform. Appl. NOTE ILLUMINA® EPIGENETIC Anal.

Wheeler, D.A., Wang, L., 2013. From human genome to cancer genome: The first decade. Genome Res. 23, 1054–1062. doi:10.1101/gr.157602.113

Wilson, B.G., Roberts, C.W.M., 2011. SWI/SNF nucleosome remodellers and cancer. Nat. Rev. Cancer 11, 481–92. doi:10.1038/nrc3068

Wing, M.R., Devaney, J.M., Joffe, M.M., Xie, D., Feldman, H.I., Dominic, E.A., Guzman, N.J., Ramezani, A., Susztak, K., Herman, J.G., Cope, L., Harmon, B., Kwabi-Addo, B., Gordish-Dressman, H., Go, A.S., He, J., Lash, J.P., Kusek, J.W., Raj, D.S., 2014. DNA methylation profile associated with rapid decline in kidney function: findings from the CRIC study. Nephrol. Dial. Transplant 29, 864–72. doi:10.1093/ndt/gft537

Wishart, D.S., 2015. Is Cancer a Genetic Disease or a Metabolic Disease? EBioMedicine 2, 478–9. doi:10.1016/j.ebiom.2015.05.022

Withanage, K., Nakagawa, K., Ikeda, M., Kurihara, H., Kudo, T., Yang, Z., Sakane, A., Sasaki, T., Hata, Y., 2012. Expression of RASSF6 in kidney and the implication of RASSF6 and the Hippo pathway in the sorbitol-induced apoptosis in renal proximal tubular epithelial cells. J. Biochem. 152, 111–119. doi:10.1093/jb/mvs056

Wolff, E.M., Chihara, Y., Pan, F., Weisenberger, D.J., Siegmund, K.D., Sugano, K., Kawashima, K., Laird, P.W., Jones, P.A., Liang, G., 2010. Unique DNA methylation patterns distinguish noninvasive and invasive urothelial cancers and establish an epigenetic field defect in premalignant tissue. Cancer Res. 70, 8169–8178. doi:10.1158/0008-5472.CAN-10-1335

World Medical Association, 1974. Declaration of Helsinki. Lance 353, 1418–1419.

Xiong, Z., Laird, P.W., 1997. COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res. 25, 2532–2534. doi:10.1093/nar/25.12.2532

Xu, B., Di, J., Wang, Z., Han, X., Li, Z., Luo, X., Zheng, Q., 2013. Quantitative analysis of RASSF1A promoter methylation in hepatocellular carcinoma and its prognostic implications. Biochem. Biophys. Res. Commun. 438, 324–328. doi:10.1016/j.bbrc.2013.07.070

Yabuta, N., Nojima, H., 2013. Hippo in Cell Cycle and Mitosis, in: The Hippo Signaling Pathway and Cancer. Springer New York, New York, NY, pp. 199–221. doi:10.1007/978-1-4614-6220-0_10

Yan, J., Xu, L., Crawford, G., Wang, Z., Burgess, S.M., 2006. The forkhead transcription factor FoxI1 remains bound to condensed mitotic chromosomes and stably remodels chromatin structure. Mol. Cell. Biol. 26, 155–68. doi:10.1128/MCB.26.1.155-168.2006

Yan, Y., Liu, L., Zhou, J., Li, L., Li, Y., Chen, M., Wang, L., He, W., Guan, X., Zu, X., Qi, L., 2015. Clinicopathologic characteristics and prognostic factors of sarcomatoid renal cell carcinoma. J. Cancer Res. Clin. Oncol. 141, 345–352. doi:10.1007/s00432-014-1740-1

Yang, L., Hu, H.-M., Zielinska-Kwiatkowska, A., Chansky, H.A., 2010. FOXO1 is a direct target of EWS-Fli1 oncogenic fusion protein in Ewing's sarcoma cells. Biochem. Biophys. Res. Commun. 402, 129–34. doi:10.1016/j.bbrc.2010.09.129

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., et al. 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell 26, 577–90. doi:10.1016/j.ccr.2014.07.028

Yang, X.J., Tan, M.-H., Kim, H.L., Ditlev, J.A., Betten, M.W., Png, C.E., Kort, E.J., Futami, K., Furge, K.A., Takahashi, M., Kanayama, H.-O., Tan, P.H., Teh, B.S., Luan, C., Wang, K., Pins, M., Tretiakova, M., Anema, J., Kahnoski, R., Nicol, T., Stadler, W., Vogelzang, N.G., Amato, R., Seligson, D., Figlin, R., Belldegrun, A., Rogers, C.G., Teh, B.T., 2005. A molecular classification of papillary renal cell carcinoma. Cancer Res. doi:10.1158/0008-5472.CAN-05-0533

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134. doi:10.1186/1471-2105-13-134

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., Carter, S.L., Getz, G., Stemke-Hale, K., Mills, G.B., Verhaak, R.G.W., 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. 4. doi:10.1038/ncomms3612

Youssef, Y.M., White, N.M. a, Grigull, J., Krizova, A., Samy, C., Mejia-Guerrero, S., Evans, A., Yousef, G.M., 2011. Accurate molecular classification of kidney cancer subtypes using MicroRNA signature. Eur. Urol. 59, 721–730. doi:10.1016/j.eururo.2011.01.004

Yu, M., Carter, K.T., Makar, K.W., Vickers, K., Ulrich, C.M., Schoen, R.E., Brenner, D., Markowitz, S.D., Grady, W.M., 2015. MethyLight Droplet Digital PCR for Detection and Absolute Quantification of Infrequently Methylated Alleles. Epigenetics 00–00.

doi:10.1080/15592294.2015.1068490

Yusenko, M. V, 2010. Molecular pathology of renal oncocytoma: a review. Int. J. Urol. 17, 602–12. doi:10.1111/j.1442-2042.2010.02574.x

Yusenko, M. V, Kuiper, R.P., Boethe, T., Ljungberg, B., van Kessel, A.G., Kovacs, G., 2009. High-resolution DNA copy number and gene expression analyses distinguish chromophobe renal cell carcinomas and renal oncocytomas. BMC Cancer 9, 152. doi:10.1186/1471-2407-9-152

Zabarovsky, E.R., Lerman, M.I., Minna, J.D., 2002. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. Oncogene 21, 6915–35. doi:10.1038/sj.onc.1205835

Zhang, H., Angelopoulos, N., Xu, Y., Grothey, A., Nunes, J., Stebbing, J., Giamas, G., 2015. Proteomic profile of KSR1-regulated signalling in response to genotoxic agents in breast cancer. Breast Cancer Res. Treat. 151, 555–68. doi:10.1007/s10549-015-3443-y

Zhang, J., Zhang, J., Cui, X., Yang, Y., Li, M., Qu, J., Li, J., Wang, J., 2015. FoxM1: a novel tumor biomarker of lung cancer. Int. J. Clin. Exp. Med. 8, 3136–40.

8.1.1 Internet references

1000genomes.org, (2015). Home | 1000 Genomes. [online] Available at: http://www.1000genomes.org [Accessed 26 Aug. 2015].

Abcam.com, (2015). DNA Methylation guide by Abcam. [online] Available at: http://www.abcam.com/epigenetics/dna-methylation-a-guide [Accessed 4 Sep. 2015].

Activemotif.com, (2015). Bisulfite Conversion Kit for DNA Methylation Analysis. [online] Available at: http://www.activemotif.com/catalog/176/Bisulfite-conversion-kit [Accessed 4 Sep. 2015].

Activemotif.com, (2015). MeDIP (Methylated DNA Immunoprecipitation) for enrichment of methylated cytosines. [online] Available at: http://www.activemotif.com/catalog/736/medip [Accessed 4 Sep. 2015].

Bio-rad.com, (2015). Droplet Digital™ PCR (ddPCR™) Technology | Applications & Technologies | Bio-Rad. [online] Available at: http://www.bio-rad.com/en-uk/applications-technologies/droplet-digital-pcr-ddpcr-technology [Accessed 6 Sep. 2015].

Cancer Research UK, (2015). Cancer incidence by ethnicity. [online] Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/ethnicity [Accessed 13 Aug. 2015].

Cancer Research UK, (2015). Kidney cancer incidence statistics. [online] Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/kidney-cancer/incidence#heading-Two [Accessed 13 Sep. 2015].

Cancer.sanger.ac.uk, (2015). COSMIC: Catalogue of Somatic Mutations in Cancer - Home Page. [online] Available at: http://Cancer.sanger.ac.uk [Accessed 26 Aug. 2015].

Cancer.sanger.ac.uk, (2015). COSMIC: Gene analysis - NF2. [online] Available at: http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=NF2&ln1=NF2&start=1&end=596&coords=AA%3AAA&sn=&ss=&hn=&sh=&stat=1&id=2898# [Accessed 10 Sep. 2015].

Cbioportal.org, (2015). cBioPortal for Cancer Genomics. [online] Available at: http://www.cbioportal.org [Accessed 15 Aug. 2015].

Code.google.com, (2015). bedtools - bedtools: a flexible suite of utilities for comparing genomic features. - Google Project Hosting. [online] Available at: https://code.google.com/p/bedtools/ [Accessed 26 Aug. 2015].

Diabetes.co.uk, (2015). Diabetes and Ethnicity. [online] Available at: http://www.diabetes.co.uk/diabetes-and-ethnicity.html [Accessed 12 Aug. 2015].

Encodeproject.org, (2015). ENCODE: Encyclopedia of DNA Elements – ENCODE. [online] Available at: http://Encodeproject.org [Accessed 9 Aug. 2015].

Ensembl.org, (2015). Ensembl Genome Browser. [online] Available at: http://Ensembl.org [Accessed 26 Aug. 2015].

Evs.gs.washington.edu, (2015). Exome Variant Server. [online] Available at: http://Evs.gs.washington.edu [Accessed 26 Aug. 2015].

Exac.broadinstitute.org, (2015). ExAC Browser. [online] Available at: http://exac.broadinstitute.org [Accessed 20 Sep. 2015].

Genecards Database, G. (2015). GeneCards - Human Genes | Gene Database | Gene Search. [online] Genecards.org. Available at: http://Genecards.org [Accessed 26 Aug. 2015].

Genecards.org, (2015). TP53BP2 Gene - GeneCards | ASPP2 Protein | ASPP2 Antibody. [online] Available at: http://www.genecards.org/cgi-bin/carddisp.pl?gene=TP53BP2 [Accessed 8 Sep. 2015].

Genome-euro.ucsc.edu, (2015). UCSC Genome Browser Home. [online] Available at: http://Genome-euro.ucsc.edu [Accessed 26 Aug. 2015].

Genome.jp, (2015). KEGG PATHWAY: p53 signaling pathway - Homo sapiens (human). [online] Available at: http://www.genome.jp/kegg-bin/show_pathway?hsa04115 [Accessed 13 Sep. 2015].

Genomics.agilent.com, (2015). SureSelect - How it Works. [online] Available at: http://www.genomics.agilent.com/article.jsp?pageId=3083 [Accessed 19 Sep. 2015].

Genomicsengland.co.uk, (2015). The 100,000 Genomes Project | Genomics England. [online] Available at: http://www.genomicsengland.co.uk/the-100000-genomes-project/ [Accessed 9 Aug. 2015].

Ijgvd.megabank.tohoku.ac.jp, (2015). Integrative Japanese Genome Variation Database [iJGVD] | Division of Biomedical Information Analysis, Department of Integrative Genomics. [online] Available at: http://ijgvd.megabank.tohoku.ac.jp [Accessed 26 Aug. 2015].

Illumina.com, (2015). Infinium HumanMethylation450 BeadChip Kit. [online] Available at: http://www.illumina.com/products/methylation_450_beadchip_kits.html [Accessed 10 Oct. 2012].

Ingenuity, (2015). Ingenuity IPA - Integrate and understand complex 'omics data. [online] Available at: http://www.ingenuity.com/products/ipa [Accessed 10 Oct. 2012].

King, R. (2000). Cancer Biology. 2nd ed. Pearson Prentice Hall.

Macmillan.org.uk, (2015). Key statistics - Macmillan Cancer Support. [online] Available at: http://www.macmillan.org.uk/Aboutus/Ouresearchandevaluation/Researchandevaluation/Keystatistics.aspx [Accessed 13 Sep. 2015].

National Cancer Institute, (2015). BRCA1 & BRCA2: Cancer Risk & Genetic Testing. [online] Available at: http://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet [Accessed 13 Aug. 2015].

National Cancer Institute, (2015). What Is Cancer?. [online] Available at: http://www.cancer.gov/about-cancer/what-is-cancer [Accessed 13 Sep. 2015].

Nature.com, (2015). A subway map of cancer pathways. [online] Available at: http://www.nature.com/nrc/poster/subpathways/index.html [Accessed 13 Sep. 2015].

Ncbi.nlm.nih.gov, (2015). JUND jun D proto-oncogene [Homo sapiens (human)] - Gene - NCBI. [online] Available at: http://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=3727 [Accessed 7 Sep. 2015].

Nhs.uk, (2015). Kidney cancer - Treatment - NHS Choices. [online] Available at: http://www.nhs.uk/conditions/cancer-of-the-kidney/pages/treatment.aspx [Accessed 22 Sep. 2015].

Pierorazio, P. (2014). Brady Urology at Johns Hopkins Hospital: Oncocytoma: A Benign Kidney Tumor Often Confused for Cancer. [online] Bradyurology.blogspot.co.uk. Available at: http://bradyurology.blogspot.co.uk/2014/11/oncocytoma-benign-kidney-tumor-often.html [Accessed 1 Sep. 2015].

Snp.ims.u-tokyo.ac.jp, (2015). JSNPs Home. [online] Available at: http://Snp.ims.u-tokyo.ac.jp [Accessed 26 Aug. 2015].

String-db.org, (2015). STRING: functional protein association networks. [online] Available at: http://string-db.org [Accessed 26 Aug. 2015].

Tcga-data.nci.nih.gov, (2015). The Cancer Genome Atlas - Data Portal. [online] Available at: https://tcga-data.nci.nih.gov/tcga/ [Accessed 26 Aug. 2015].

Thermofisher.com, (2015). Ion PGM™ System for Next-Generation Sequencing | Thermo Fisher Scientific. [online] Available at: https://www.thermofisher.com/uk/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-pgm-system-for-next-generation-sequencing.html [Accessed 19 Sep. 2015].

Trusight Sequencing Panels. *Illumina.com*. N.p., 2016. Web. 13 July 2016.
 Who.int, (2015). WHO | Cancer. [online] Available at: http://www.who.int/cancer/en/ [Accessed 13 Sep. 2015].

Slater, A.A., Alokail, M., Gentle, D., Yao, M., Kovacs, G., Maher, E.R., Latif, F., 2013. DNA methylation profiling distinguishes histological subtypes of renal cell carcinoma. Epigenetics 8, 252–67. doi:10.4161/epi.23817